BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/60, 9/88, 1/21, C12Q 1/68 | A2

(11) International Publication Number: **WO 99/49058**

(43) International Publication Date: 30 September 1999 (30.09.99)

(21) International Application Number: PCT/US99/06046

(22) International Filing Date: 19 March 1999 (19.03.99)

(30) Priority Data:
60/079,386 26 March 1998 (26.03.98) US

(71) Applicant *(for all designated States except US)*: E.I. DU PONT DE NEMOURS AND COMPANY [US/US]; 1007 Market Street, Wilmington, DE 19898 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: VOLLMER, Steven, J. [US/US]; 1019 Tweedbrook Road, Wilmington, DE 19810 (US). FALCO, Saverio, Carl [US/US]; 1902 Millers Road, Arden, DE 19810 (US). BROGLIE, Richard, M. [US/US]; 520 Port Royal Court, Landenberg, PA 19350 (US). BRYAN, Gregory, T. [NZ/US]; 1215 Spruce Street, Wilmington, DE 19805 (US). CAHOON, Rebecca, E. [US/US]; 2331 West 18th Street, Wilmington, DE 19806 (US). RAFALSKI, J., Antoni [US/US]; 2028 Longcome Drive, Wilmington, DE 19810 (US).

(74) Agent: MAJARIAN, William, R.; E.I. du Pont de Nemours and Company, Legal Patent Records Center, 1007 Market Street, Wilmington, DE 19898 (US).

(81) Designated States: AL, AU, BA, BB, BG, BR, CA, CN, CU, CZ, EE, GD, GE, HR, HU, ID, IL, IN, IS, JP, KP, KR, LC, LK, LR, LT, LV, MG, MK, MN, MX, NO, NZ, PL, RO, SG, SI, SK, SL, TR, TT, UA, US, UZ, VN, YU, ZA, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: TRYPTOPHAN BIOSYNTHETIC ENZYMES

(57) Abstract

This invention relates to an isolated nucleic acid fragment encoding a tryptophan biosynthetic enzyme. The invention also relates to the construction of a chimeric gene encoding all or a portion of the tryptophan biosynthetic enzyme, in sense or antisense orientation, wherein expression of the chimeric gene results in production of altered levels of the tryptophan biosynthetic enzyme in a transformed host cell.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

TITLE

TRYPTOPHAN BIOSYNTHETIC ENZYMES

This application claims the benefit of U.S. Provisional Application No. 60/079,386, filed March 26, 1998.

FIELD OF THE INVENTION

This invention is in the field of plant molecular biology. More specifically, this invention pertains to nucleic acid fragments encoding enzymes involved in the biosynthesis of tyrotophan in plants and seeds.

BACKGROUND OF THE INVENTION

Many vertebrates, including man, lack the ability to manufacture a number of amino acids and therefore require these amino acids preformed in their diet. These are called essential amino acids. Plants are able to synthesize all twenty amino acids and serve as the ultimate source of the essential amino acids for humans and animals. Thus, the ability to manipulate the production and accumulation of the essential amino acids in plants is of considerable importance and value. Furthermore, the inability of animals to synthesize these amino acids provides a useful distinction between animal and plant cellular metabolism. This can be exploited for the discovery of herbicidal chemical compounds that target enzymes in the plant biosynthetic pathways of the essential amino acids and thus have low toxicity to animals.

Tryptophan is an essential amino acid. In plants, the biosynthesis of tryptophan from chorismic acid (see Figure 1) requires five enzymatic steps catalyzed by anthranilate synthase (EC 4.1.3.27), anthranilate phosphoribosyl-transferase (EC 2.4.2.18), phosphoribosylanthranilate isomerase (EC 5.3.1.24), indole-3-glycerol phosphate synthase (EC 4.1.1.48) and tryptophan synthase (EC 4.2.1.20). The tryptophan pathway leads to the biosynthesis of many secondary metabolites including the hormone indole-3-acetic acid, antimicrobial phytoalexins, alkaloids and glucosinolates. The first enzyme in the tryptophan pathway from chorismate is anthranilate synthase. This enzyme is subject to feedback inhibition by tryptophan and is composed of two subunits. Even though the anthranilate synthase alpha subunits are encoded by duplicate genes, in *Arabidopsis* and *Ruta graveolens* it has been shown that only one gene is induced in response to wounding or pathogen attack (Bohlmann, J. et al. (1995) *Plant J.* 7:491-501 and Niyogi, K. K. and Fink, G. R. (1992) *Plant Cell* 4:721-733). Glutamine- and ammonia-dependent anthranilate synthase activities copurify in *Ruta graveolens*. The glutamine-dependent reaction of one of the alpha subunits requires a 60 to 65 kDa anthranilate synthase beta subunit (Bohlmann, J. et al. (1995) *Plant J.* 7:491-501) The beta-subunit is a highly asymmetric dimer with an apparent molecular weight of 200,000. Cleavage of the purified subunit with elastase, trypsin, or chymotrypsin results in fragments which retain enzyme activity. Elastase digestion results in a 30 kDa fragment and a 56 kDa fragment. The first fragment behaves as a monomer and interacts with free alpha subunit to produce the glutamine-dependent anthranilate synthase activity.

The second fragment behaves as an asymmetric dimer and has N-(5'-phosphoribosyl) anthranilate isomerase and indole-3-glycerol phosphate synthase activity (Walker, M. S. and DeMoss, J. A. (1983) *J Biol Chem 258*:3571-3575).

The two final reactions in tryptophan biosynthesis are catalyzed by tryptophan synthase. The 29 kDa alpha subunit is a bifunctional enzyme which cleaves indole-3-glycerol phosphate to produce indole and glyceraldehyde-3-phosphate. The beta subunit joins indole with serine to form tryptophan. Either subunit alone is enzymatically active, but the rate of the reaction and affinity for the substrates increases when the subunits are forming a tetramer composed of two alpha subunits and two beta subunits (Radwanski, E. R. (1995) *Mol Gen Genet 248*:657-667).

Few of the genes encoding enzymes from the tryptophan pathway in corn, soybeans, rice and wheat, have been isolated and sequenced. For example, no corn, soybean, rice or wheat genes have been reported for anthranilate synthase alpha or beta subunits and no soybean, rice or wheat genes have been reported for tryptophan synthase alpha subunit. A corn gene encoding tryptophan synthase has been identified and the instant invention describes a new corn tryptophan synthase homolog. Accordingly, the availability of nucleic acid sequences encoding all or a portion of these enzymes will facilitate studies to better understand cellular biosynthetic pathways, provide genetic tools for the manipulation of those pathways, provide a means to evaluate chemical compounds for their ability to inhibit the enzymatic activity of the compounds included in this application.

SUMMARY OF THE INVENTION

The instant invention relates to isolated nucleic acid fragments encoding tryptophan biosynthetic enzymes. Specifically, this invention concerns an isolated nucleic acid fragment encoding an anthranilate synthase alpha subunit, an anthranilate synthase beta subunit or a tryptophan synthase alpha subunit. In addition, this invention relates to a nucleic acid fragment that is complementary to the nucleic acid fragment encoding anthranilate synthase alpha subunit, anthranilate synthase beta subunit or tryptophan synthase alpha subunit.

An additional embodiment of the instant invention pertains to a polypeptide encoding all or a substantial portion of a tryptophan biosynthetic enzyme selected from the group consisting of anthranilate synthase alpha subunit, anthranilate synthase beta subunit and tryptophan synthase alpha subunit.

In another embodiment, the instant invention relates to a chimeric gene encoding an anthranilate synthase alpha subunit, an anthranilate synthase beta subunit or a tryptophan synthase alpha subunit, or to a chimeric gene that comprises a nucleic acid fragment that is complementary to a nucleic acid fragment encoding an anthranilate synthase alpha subunit, an anthranilate synthase beta subunit or a tryptophan synthase alpha subunit, operably linked to suitable regulatory sequences, wherein expression of the chimeric gene results in

production of levels of the encoded protein in a transformed host cell that is altered (i.e., increased or decreased) from the level produced in an untransformed host cell.

In a further embodiment, the instant invention concerns a transformed host cell comprising in its genome a chimeric gene encoding an anthranilate synthase alpha subunit, an anthranilate synthase beta subunit or a tryptophan synthase alpha subunit, operably linked to suitable regulatory sequences. Expression of the chimeric gene results in production of altered levels of the encoded protein in the transformed host cell. The transformed host cell can be of eukaryotic or prokaryotic origin, and include cells derived from higher plants and microorganisms. The invention also includes transformed plants that arise from transformed host cells of higher plants, and seeds derived from such transformed plants.

An additional embodiment of the instant invention concerns a method of altering the level of expression of an anthranilate synthase alpha subunit, an anthranilate synthase beta subunit or a tryptophan synthase alpha subunit in a transformed host cell comprising: a) transforming a host cell with a chimeric gene comprising a nucleic acid fragment encoding an anthranilate synthase alpha subunit, an anthranilate synthase beta subunit or a tryptophan synthase alpha subunit; and b) growing the transformed host cell under conditions that are suitable for expression of the chimeric gene wherein expression of the chimeric gene results in production of altered levels of anthranilate synthase alpha subunit, anthranilate synthase beta subunit or tryptophan synthase alpha subunit in the transformed host cell.

An addition embodiment of the instant invention concerns a method for obtaining a nucleic acid fragment encoding all or a substantial portion of an amino acid sequence encoding an anthranilate synthase alpha subunit, an anthranilate synthase beta subunit or a tryptophan synthase alpha subunit.

A further embodiment of the instant invention is a method for evaluating at least one compound for its ability to inhibit the activity of an anthranilate synthase alpha subunit, an anthranilate synthase beta subunit or a tryptophan synthase alpha subunit, the method comprising the steps of: (a) transforming a host cell with a chimeric gene comprising a nucleic acid fragment encoding an anthranilate synthase alpha subunit, an anthranilate synthase beta subunit or a tryptophan synthase alpha subunit, operably linked to suitable regulatory sequences; (b) growing the transformed host cell under conditions that are suitable for expression of the chimeric gene wherein expression of the chimeric gene results in production of anthranilate synthase alpha subunit, anthranilate synthase beta subunit or tryptophan synthase alpha subunit in the transformed host cell; (c) optionally purifying the anthranilate synthase alpha subunit, the anthranilate synthase beta subunit or the tryptophan synthase alpha subunit expressed by the transformed host cell; (d) treating the anthranilate synthase alpha subunit, the anthranilate synthase beta subunit or the tryptophan synthase alpha subunit with a compound to be tested; and (e) comparing the activity of the anthranilate synthase alpha subunit, the anthranilate synthase beta subunit or the tryptophan

synthase alpha subunit that has been treated with a test compound to the activity of an untreated anthranilate synthase alpha subunit, anthranilate synthase beta subunit or tryptophan synthase alpha subunit, thereby selecting compounds with potential for inhibitory activity.

BRIEF DESCRIPTION OF THE DRAWINGS AND SEQUENCE DESCRIPTIONS

The invention can be more fully understood from the following detailed description and the accompanying drawings and Sequence Listing which form a part of this application.

Figure 1 depicts biochemical pathway for the production of tryptophan from chorismate.

Figure 2 shows an alignment of the amino acid sequences from *Ruta graveolens* anthranilate synthase alpha subunit (SEQ ID NO:39) and the instant corn anthranilate synthase alpha subunit (contig of cde1c.pk004.g2, p0125.czabf83r, p0005.cbmew77r, p0096.cnamt32r, p0032.crcas26r, p0031.ccmar07r, p0015.cdpfk12r, p0128.cpibl07r and cta1n.pk0051.g3; SEQ ID NO:2). Amino acid which are identical among both sequences are indicated with an asterisk (*) above the alignment. Dashes are used by the program to maximize alignment of the sequences.

Figure 3 shows an alignment of the amino acid sequences from *Arabidopsis thaliana* anthranilate synthase beta subunit (SEQ ID NO:40), the instant corn anthranilate synthase beta subunit (contig of p0126.cnlcu75r, cen3n.pk0212.h6, cco1n.pk0038.b1, csi1n.pk0017.d4, p0031.ccmai08r and ctn1c.pk001.l10; SEQ ID NO:20) and the instant rice anthranilate synthase beta subunit (contig of rds2c.pk004.i9, rls24.pk0002.e2 and rl0n.pk113.b4; SEQ ID NO:22). Amino acid which are identical among all sequences are indicated with an asterisk (*) above the alignment. Dashes are used by the program to maximize alignment of the sequences.

Figure 4 shows an alignment of the amino acid sequences from *Arabidopsis thaliana* tryptophan synthase alpha subunit (SEQ ID NO:41), the instant corn tryptophan synthase alpha subunit (contig of chp2.pk0020.e2 and lkr.pk0013.g1; SEQ ID NO:26), the instant corn tryptophan synthase alpha subunit (cr1n.pk0052.b8; SEQ ID NO:28) and the instant rice tryptophan synthase alpha subunit (rr1.pk0038.h6; SEQ ID NO:32). Amino acid which are identical among all sequences are indicated with an asterisk (*) above the alignment. Dashes are used by the program to maximize alignment of the sequences.

The following sequence descriptions and Sequence Listing attached hereto comply with the rules governing nucleotide and/or amino acid sequence disclosures in patent applications as set forth in 37 C.F.R. §1.821-1.825.

SEQ ID NO:1 is the nucleotide sequence comprising the contig assembled from a portion of the cDNA inserts in clones cde1c.pk004.g2, p0125.czabf83r, p0005.cbmew77r, p0096.cnamt32r, p0032.crcas26r, p0031.ccmar07r, p0015.cdpfk12r and p0128.cpibl07r and

the entire cDNA insert in clone cta1n.pk0051.g3 encoding an entire corn anthranilate synthase alpha subunit.

SEQ ID NO:2 is the deduced amino acid sequence of an entire corn anthranilate synthase alpha subunit derived from the nucleotide sequence of SEQ ID NO:1.

SEQ ID NO:3 is the nucleotide sequence comprising the contig assembled from a portion of the cDNA inserts in clones p0115.clsmj70r, cbn2.pk0049.d2, cpl1c.pk008.m4 and p0102.ceraq01r and the entire cDNA insert in clone ceb3.pk0011.c12 encoding a portion of a corn anthranilate synthase alpha subunit.

SEQ ID NO:4 is the deduced amino acid sequence of a portion of a corn anthranilate synthase alpha subunit derived from the nucleotide sequence of SEQ ID NO:3.

SEQ ID NO:5 is the nucleotide sequence comprising a portion of the cDNA insert in clone rca1n.pk004.p22 encoding a portion of a rice anthranilate synthase alpha subunit.

SEQ ID NO:6 is the deduced amino acid sequence of a portion of a rice anthranilate synthase alpha subunit derived from the nucleotide sequence of SEQ ID NO:5.

SEQ ID NO:7 is the nucleotide sequence comprising a portion of the cDNA insert in clone sdp3c.pk019.c6 encoding a portion of a soybean anthranilate synthase alpha subunit.

SEQ ID NO:8 is the deduced amino acid sequence of a portion of a soybean anthranilate synthase alpha subunit derived from the nucleotide sequence of SEQ ID NO:7.

SEQ ID NO:9 is the nucleotide sequence comprising a portion of the cDNA insert in clone sfl1.pk129.d10 encoding a portion of a soybean anthranilate synthase alpha subunit.

SEQ ID NO:10 is the deduced amino acid sequence of a portion of a soybean anthranilate synthase alpha subunit derived from the nucleotide sequence of SEQ ID NO:9.

SEQ ID NO:11 is the nucleotide sequence comprising a portion of the 5' terminus of the cDNA insert in clone wle1n.pk0075.b4 encoding a portion of a wheat anthranilate synthase alpha subunit.

SEQ ID NO:12 is the deduced amino acid sequence of a portion of a wheat anthranilate synthase alpha subunit derived from the nucleotide sequence of SEQ ID NO:11.

SEQ ID NO:13 is the nucleotide sequence comprising a portion of the 3' terminus of the cDNA insert in clone wle1n.pk0075.b4 encoding a portion of a wheat anthranilate synthase alpha subunit.

SEQ ID NO:14 is the deduced amino acid sequence of a portion of a wheat anthranilate synthase alpha subunit derived from the nucleotide sequence of SEQ ID NO:13.

SEQ ID NO:15 is the nucleotide sequence comprising a portion of the cDNA insert in clone wlm96.pk026.j1 encoding a portion of a wheat anthranilate synthase alpha subunit.

SEQ ID NO:16 is the deduced amino acid sequence of a portion of a wheat anthranilate synthase alpha subunit derived from the nucleotide sequence of SEQ ID NO:15.

SEQ ID NO:17 is the nucleotide sequence comprising the contig assembled from a portion of the cDNA insert in clones cen3n.pk0210.c3, cepe7.pk0012.g2 and cen3n.pk0047.g5 encoding a portion of a corn anthranilate synthase beta subunit.

SEQ ID NO:18 is the deduced amino acid sequence of a portion of a corn anthranilate synthase beta subunit derived from the nucleotide sequence of SEQ ID NO:17.

SEQ ID NO:19 is the nucleotide sequence comprising the contig assembled from a portion of the cDNA inserts in clones p0126.cnlcu75r, cen3n.pk0212.h6, cco1n.pk0038.b1, csi1n.pk0017.d4, p0031.ccmai08r and ctn1c.pk001.l10 encoding an entire corn anthranilate synthase beta subunit.

SEQ ID NO:20 is the deduced amino acid sequence of an entire corn anthranilate synthase beta subunit derived from the nucleotide sequence of SEQ ID NO:19.

SEQ ID NO:21 is the nucleotide sequence comprising the contig assembled from a portion of the cDNA inserts in clones rds2c.pk004.i9, rls24.pk0002.e2 and rl0n.pk113.b4 encoding a a substantial portion of a rice anthranilate synthase beta subunit.

SEQ ID NO:22 is the deduced amino acid sequence of a substantial portion of a rice anthranilate synthase beta subunit derived from the nucleotide sequence of SEQ ID NO:21.

SEQ ID NO:23 is the nucleotide sequence comprising a portion of the cDNA insert in clone wlm96.pk045.c11 encoding a portion of a wheat anthranilate synthase beta subunit.

SEQ ID NO:24 is the deduced amino acid sequence of a portion of a wheat anthranilate synthase beta subunit derived from the nucleotide sequence of SEQ ID NO:23.

SEQ ID NO:25 is the nucleotide sequence comprising the contig assembled from a portion of the cDNA inserts in clones chp2.pk0020.e2 and lkr.pk0013.g1 encoding an entire corn tryptophan synthase alpha subunit.

SEQ ID NO:26 is the deduced amino acid sequence of an entire corn tryptophan synthase alpha subunit derived from the nucleotide sequence of SEQ ID NO:25.

SEQ ID NO:27 is the nucleotide sequence comprising a portion of the cDNA insert in clone cr1n.pk0052.b8 encoding an almost entire corn tryptophan synthase alpha subunit.

SEQ ID NO:28 is the deduced amino acid sequence of an almost entire corn tryptophan synthase alpha subunit derived from the nucleotide sequence of SEQ ID NO:27.

SEQ ID NO:29 is the nucleotide sequence comprising the contig assembled from a portion of the cDNA inserts in clones res1c.pk004.c4 and rr1.pk0071.f2 encoding a portion of a rice tryptophan synthase alpha subunit.

SEQ ID NO:30 is the deduced amino acid sequence of a portion of a rice tryptophan synthase alpha subunit derived from the nucleotide sequence of SEQ ID NO:29.

SEQ ID NO:31 is the nucleotide sequence comprising a portion of the cDNA insert in clone rr1.pk0038.h6 encoding an entire rice tryptophan synthase alpha subunit.

SEQ ID NO:32 is the deduced amino acid sequence of an entire rice tryptophan synthase alpha subunit derived from the nucleotide sequence of SEQ ID NO:31.

SEQ ID NO:33 is the nucleotide sequence comprising the contig assembled from a portion of the cDNA inserts in clones src3c.pk004.c19and sls2c.pk025.d11 encoding a portion of a soybean tryptophan synthase alpha subunit.

SEQ ID NO:34 is the deduced amino acid sequence of a portion of a soybean tryptophan synthase alpha subunit derived from the nucleotide sequence of SEQ ID NO:33.

SEQ ID NO:35 is the nucleotide sequence comprising the contig assembled from a portion of the cDNA inserts in clones wlm0.pk0028.f4 and wlm96.pk041.h3 encoding a portion of a wheat tryptophan synthase alpha subunit.

SEQ ID NO:36 is the deduced amino acid sequence of a portion of a wheat tryptophan synthase alpha subunit derived from the nucleotide sequence of SEQ ID NO:35.

SEQ ID NO:37 is the nucleotide sequence comprising the contig assembled from a portion of the cDNA inserts in clones wl1n.pk0109.d5 and wl1n.pk0110.c1 encoding a portion of a wheat tryptophan synthase alpha subunit.

SEQ ID NO:38 is the deduced amino acid sequence of a portion of a wheat tryptophan synthase alpha subunit derived from the nucleotide sequence of SEQ ID NO:37.

SEQ ID NO:39 is the amino acid sequence of a *Ruta graveolens* anthranilate synthase alpha subunit having an NCBI General Identifier No. 960289.

SEQ ID NO:40 is the amino acid sequence of a *Arabidopsis thaliana* anthranilate synthase beta subunit having an NCBI General Identifier No. 541849.

SEQ ID NO:41 is the amino acid sequence of a *Arabidopsis thaliana* tryptophan synthase alpha subunit having an NCBI General Identifier No. 2129755.

The Sequence Listing contains the one letter code for nucleotide sequence characters and the three letter codes for amino acids as defined in conformity with the IUPAC-IUBMB standards described in *Nucleic Acids Research 13*:3021-3030 (1985) and in the *Biochemical Journal 219 (No. 2)*:345-373 (1984) which are herein incorporated by reference. The symbols and format used for nucleotide and amino acid sequence data comply with the rules set forth in 37 C.F.R. §1.822.

DETAILED DESCRIPTION OF THE INVENTION

In the context of this disclosure, a number of terms shall be utilized. As used herein, an "isolated nucleic acid fragment" is a polymer of RNA or DNA that is single- or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases. An isolated nucleic acid fragment in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA or synthetic DNA. As used herein, "contig" refers to an assemblage of overlapping nucleic acid sequences to form one contiguous nucleotide sequence. For example, several DNA sequences can be compared and aligned to identify common or overlapping regions. The individual sequences can then be assembled into a single contiguous nucleotide sequence.

As used herein, "substantially similar" refers to nucleic acid fragments wherein changes in one or more nucleotide bases results in substitution of one or more amino acids, but do not affect the functional properties of the protein encoded by the DNA sequence. "Substantially similar" also refers to nucleic acid fragments wherein changes in one or more nucleotide bases does not affect the ability of the nucleic acid fragment to mediate alteration

of gene expression by antisense or co-suppression technology. "Substantially similar" also refers to modifications of the nucleic acid fragments of the instant invention such as deletion or insertion of one or more nucleotides that do not substantially affect the functional properties of the resulting transcript vis-à-vis the ability to mediate alteration of gene expression by antisense or co-suppression technology or alteration of the functional properties of the resulting protein molecule. It is therefore understood that the invention encompasses more than the specific exemplary sequences.

For example, it is well known in the art that antisense suppression and co-suppression of gene expression may be accomplished using nucleic acid fragments representing less than the entire coding region of a gene, and by nucleic acid fragments that do not share 100% sequence identity with the gene to be suppressed. Moreover, alterations in a gene which result in the production of a chemically equivalent amino acid at a given site, but do not effect the functional properties of the encoded protein, are well known in the art. Thus, a codon for the amino acid alanine, a hydrophobic amino acid, may be substituted by a codon encoding another less hydrophobic residue, such as glycine, or a more hydrophobic residue, such as valine, leucine, or isoleucine. Similarly, changes which result in substitution of one negatively charged residue for another, such as aspartic acid for glutamic acid, or one positively charged residue for another, such as lysine for arginine, can also be expected to produce a functionally equivalent product. Nucleotide changes which result in alteration of the N-terminal and C-terminal portions of the protein molecule would also not be expected to alter the activity of the protein. Each of the proposed modifications is well within the routine skill in the art, as is determination of retention of biological activity of the encoded products. Moreover, the skilled artisan recognizes that substantially similar nucleic acid sequences encompassed by this invention are also defined by their ability to hybridize, under stringent conditions (0.1X SSC, 0.1% SDS, 65°C), with the sequences exemplified herein. Substantially similar nucleic acid fragments of the instant invention may also be characterized by the percent identity of the amino acid sequences that they encode to the amino acid sequences disclosed herein, as determined by algorithms commonly employed by those skilled in this art. Preferred are those nucleic acid fragments whose nucleotide sequences encode amino acid sequences that are greater than 80% identical to the amino acid sequences reported herein. More preferred nucleic acid fragments encode amino acid sequences that are 90% identical to the amino acid sequences reported herein. Most preferred are nucleic acid fragments that encode amino acid sequences that are 95% identical to the amino acid sequences reported herein. Sequence alignments and percent similarity calculations were performed using the Megalign program of the LASARGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins, D.G. and Sharp, P.M. (1989) *CABIOS.* 5:151-153) with the default parameters (GAP PENALTY=10, GAP

LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method were KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5.

A "substantial portion" of an amino acid or nucleotide sequence comprises enough of the amino acid sequence of a polypeptide or the nucleotide sequence of a gene to afford putative identification of that polypeptide or gene, either by manual evaluation of the sequence by one skilled in the art, or by computer-automated sequence comparison and identification using algorithms such as BLAST (Basic Local Alignment Search Tool; Altschul, S. F., et al., (1993) *J. Mol. Biol. 215*:403-410; see also www.ncbi.nlm.nih.gov/BLAST/). In general, a sequence of ten or more contiguous amino acids or thirty or more nucleotides is necessary in order to putatively identify a polypeptide or nucleic acid sequence as homologous to a known protein or gene. Moreover, with respect to nucleotide sequences, gene specific oligonucleotide probes comprising 20-30 contiguous nucleotides may be used in sequence-dependent methods of gene identification (e.g., Southern hybridization) and isolation (e.g., *in situ* hybridization of bacterial colonies or bacteriophage plaques). In addition, short oligonucleotides of 12-15 bases may be used as amplification primers in PCR in order to obtain a particular nucleic acid fragment comprising the primers. Accordingly, a "substantial portion" of a nucleotide sequence comprises enough of the sequence to afford specific identification and/or isolation of a nucleic acid fragment comprising the sequence. The instant specification teaches partial or complete amino acid and nucleotide sequences encoding one or more particular plant proteins. The skilled artisan, having the benefit of the sequences as reported herein, may now use all or a substantial portion of the disclosed sequences for purposes known to those skilled in this art. Accordingly, the instant invention comprises the complete sequences as reported in the accompanying Sequence Listing, as well as substantial portions of those sequences as defined above.

"Codon degeneracy" refers to divergence in the genetic code permitting variation of the nucleotide sequence without effecting the amino acid sequence of an encoded polypeptide. Accordingly, the instant invention relates to any nucleic acid fragment that encodes all or a substantial portion of the amino acid sequence encoding the anthranilate synthase alpha subunit, the anthranilate synthase beta subunit or the tryptophan synthase alpha subunit proteins as set forth in SEQ ID NOs:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36 and 38. The skilled artisan is well aware of the "codon-bias" exhibited by a specific host cell in usage of nucleotide codons to specify a given amino acid. Therefore, when synthesizing a gene for improved expression in a host cell, it is desirable to design the gene such that its frequency of codon usage approaches the frequency of preferred codon usage of the host cell.

"Synthetic genes" can be assembled from oligonucleotide building blocks that are chemically synthesized using procedures known to those skilled in the art. These building blocks are ligated and annealed to form gene segments which are then enzymatically

assembled to construct the entire gene. "Chemically synthesized", as related to a sequence of DNA, means that the component nucleotides were assembled *in vitro*. Manual chemical synthesis of DNA may be accomplished using well established procedures, or automated chemical synthesis can be performed using one of a number of commercially available machines. Accordingly, the genes can be tailored for optimal gene expression based on optimization of nucleotide sequence to reflect the codon bias of the host cell. The skilled artisan appreciates the likelihood of successful gene expression if codon usage is biased towards those codons favored by the host. Determination of preferred codons can be based on a survey of genes derived from the host cell where sequence information is available.

"Gene" refers to a nucleic acid fragment that expresses a specific protein, including regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence. "Native gene" refers to a gene as found in nature with its own regulatory sequences. "Chimeric gene" refers any gene that is not a native gene, comprising regulatory and coding sequences that are not found together in nature. Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature. "Endogenous gene" refers to a native gene in its natural location in the genome of an organism. A "foreign" gene refers to a gene not normally found in the host organism, but that is introduced into the host organism by gene transfer. Foreign genes can comprise native genes inserted into a non-native organism, or chimeric genes. A "transgene" is a gene that has been introduced into the genome by a transformation procedure.

"Coding sequence" refers to a DNA sequence that codes for a specific amino acid sequence. "Regulatory sequences" refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences may include promoters, translation leader sequences, introns, and polyadenylation recognition sequences.

"Promoter" refers to a DNA sequence capable of controlling the expression of a coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. The promoter sequence consists of proximal and more distal upstream elements, the latter elements often referred to as enhancers. Accordingly, an "enhancer" is a DNA sequence which can stimulate promoter activity and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue-specificity of a promoter. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic DNA segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions. Promoters which cause a

gene to be expressed in most cell types at most times are commonly referred to as "constitutive promoters". New promoters of various types useful in plant cells are constantly being discovered; numerous examples may be found in the compilation by Okamuro and Goldberg, (1989) *Biochemistry of Plants 15*:1-82. It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, DNA fragments of different lengths may have identical promoter activity.

The "translation leader sequence" refers to a DNA sequence located between the promoter sequence of a gene and the coding sequence. The translation leader sequence is present in the fully processed mRNA upstream of the translation start sequence. The translation leader sequence may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency. Examples of translation leader sequences have been described (Turner, R. and Foster, G. D. (1995) *Molecular Biotechnology 3*:225).

The "3' non-coding sequences" refer to DNA sequences located downstream of a coding sequence and include polyadenylation recognition sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor. The use of different 3' non-coding sequences is exemplified by Ingelbrecht et al., (1989) *Plant Cell 1*:671-680.

"RNA transcript" refers to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived from posttranscriptional processing of the primary transcript and is referred to as the mature RNA. "Messenger RNA (mRNA)" refers to the RNA that is without introns and that can be translated into protein by the cell. "cDNA" refers to a double-stranded DNA that is complementary to and derived from mRNA. "Sense" RNA refers to RNA transcript that includes the mRNA and so can be translated into protein by the cell. "Antisense RNA" refers to a RNA transcript that is complementary to all or part of a target primary transcript or mRNA and that blocks the expression of a target gene (U.S. Pat. No. 5,107,065, incorporated herein by reference). The complementarity of an antisense RNA may be with any part of the specific gene transcript, i.e., at the 5' non-coding sequence, 3' non-coding sequence, introns, or the coding sequence. "Functional RNA" refers to sense RNA, antisense RNA, ribozyme RNA, or other RNA that may not be translated but yet has an effect on cellular processes.

The term "operably linked" refers to the association of nucleic acid sequences on a single nucleic acid fragment so that the function of one is affected by the other. For example, a promoter is operably linked with a coding sequence when it is capable of affecting the expression of that coding sequence (i.e., that the coding sequence is under the transcriptional control of the promoter). Coding sequences can be operably linked to regulatory sequences in sense or antisense orientation.

The term "expression", as used herein, refers to the transcription and stable accumulation of sense (mRNA) or antisense RNA derived from the nucleic acid fragment of the invention. Expression may also refer to translation of mRNA into a polypeptide. "Antisense inhibition" refers to the production of antisense RNA transcripts capable of suppressing the expression of the target protein. "Overexpression" refers to the production of a gene product in transgenic organisms that exceeds levels of production in normal or non-transformed organisms. "Co-suppression" refers to the production of sense RNA transcripts capable of suppressing the expression of identical or substantially similar foreign or endogenous genes (U.S. Pat. No. 5,231,020, incorporated herein by reference).

"Altered levels" refers to the production of gene product(s) in transgenic organisms in amounts or proportions that differ from that of normal or non-transformed organisms.

"Mature" protein refers to a post-translationally processed polypeptide; i.e., one from which any pre- or propeptides present in the primary translation product have been removed. "Precursor" protein refers to the primary product of translation of mRNA; i.e., with pre- and propeptides still present. Pre- and propeptides may be but are not limited to intracellular localization signals.

A "chloroplast transit peptide" is an amino acid sequence which is translated in conjunction with a protein and directs the protein to the chloroplast or other plastid types present in the cell in which the protein is made. "Chloroplast transit sequence" refers to a nucleotide sequence that encodes a chloroplast transit peptide. A "signal peptide" is an amino acid sequence which is translated in conjunction with a protein and directs the protein to the secretory system (Chrispeels, J. J., (1991) *Ann. Rev. Plant Phys. Plant Mol. Biol. 42*:21-53). If the protein is to be directed to a vacuole, a vacuolar targeting signal (*supra*) can further be added, or if to the endoplasmic reticulum, an endoplasmic reticulum retention signal (*supra*) may be added. If the protein is to be directed to the nucleus, any signal peptide present should be removed and instead a nuclear localization signal included (Raikhel (1992) *Plant Phys.100*:1627-1632).

"Transformation" refers to the transfer of a nucleic acid fragment into the genome of a host organism, resulting in genetically stable inheritance. Host organisms containing the transformed nucleic acid fragments are referred to as "transgenic" organisms. Examples of methods of plant transformation include Agrobacterium-mediated transformation (De Blaere et al. (1987) *Meth. Enzymol. 143*:277) and particle-accelerated or "gene gun" transformation technology (Klein et al. (1987) *Nature (London) 327*:70-73; U.S. Pat. No. 4,945,050, incorporated herein by reference).

Standard recombinant DNA and molecular cloning techniques used herein are well known in the art and are described more fully in Sambrook, J., Fritsch, E.F. and Maniatis, T. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, 1989 (hereinafter "Maniatis").

Nucleic acid fragments encoding at least a portion of several tryptophan biosynthetic enzymes have been isolated and identified by comparison of random plant cDNA sequences to public databases containing nucleotide and protein sequences using the BLAST algorithms well known to those skilled in the art. Table 1 lists the proteins that are described herein, and the designation of the cDNA clones that comprise the nucleic acid fragments encoding these proteins.

TABLE 1

Tryptophan Biosynthetic Enzymes

Enzyme	Clone	Plant
Anthranilate Synthase Alpha Subunit	Contig of: cde1c.pk004.g2 p0125.czabf83r p0005.cbmew77r p0096.cnamt32r p0032.crcas26r p0031.ccmar07r p0015.cdpfk12r p0128.cpibl07r cta1n.pk0051.g3	Corn
	Contig of: p0115.clsmj70r cbn2.pk0049.d2 cpl1c.pk008.m4 p0102.ceraq01r ceb3.pk0011.c12	Corn
	rca1n.pk004.p22	Rice
	sdp3c.pk019.c6	Soybean
	sfl1.pk129.d10	Soybean
	wle1n.pk0075.b4	Wheat
	wlm96.pk026.j1	Wheat
Anthranilate Synthase Beta Subunit	Contig of: cen3n.pk0210.c3 cepe7.pk0012.g2 cen3n.pk0047.g5	Corn
	Contig of: p0126.cnlcu75r cen3n.pk0212.h6 cco1n.pk0038.b1 csi1n.pk0017.d4 p0031.ccmai08r ctn1c.pk001.l10	Corn
	Contig of: rds2c.pk004.i9 rls24.pk0002.e2 rl0n.pk113.b4	Rice
	wlm96.pk045.c11	Wheat

Enzyme	Clone	Plant
Tryptophan Synthase Alpha Subunit	Contig of: chp2.pk0020.e2 lkr.pk0013.g1	Corn
	cr1n.pk0052.b8	Corn
	Contig of: res1c.pk004.c4 rr1.pk0071.f2	Rice
	rr1.pk0038.h6	Rice
	Contig of: src3c.pk004.c19 sls2c.pk025.d11	Soybean
	Contig of: wlm0.pk0028.f4 wlm96.pk041.h3	Wheat
	Contig of: wl1n.pk0109.d5 wl1n.pk0110.c1	Wheat

The nucleic acid fragments of the instant invention may be used to isolate cDNAs and genes encoding homologous proteins from the same or other plant species. Isolation of homologous genes using sequence-dependent protocols is well known in the art. Examples of sequence-dependent protocols include, but are not limited to, methods of nucleic acid hybridization, and methods of DNA and RNA amplification as exemplified by various uses of nucleic acid amplification technologies (e.g., polymerase chain reaction, ligase chain reaction).

For example, genes encoding other anthranilate synthase alpha subunit, anthranilate synthase beta subunit or tryptophan synthase alpha subunit, either as cDNAs or genomic DNAs, could be isolated directly by using all or a portion of the instant nucleic acid fragments as DNA hybridization probes to screen libraries from any desired plant employing methodology well known to those skilled in the art. Specific oligonucleotide probes based upon the instant nucleic acid sequences can be designed and synthesized by methods known in the art (Maniatis). Moreover, the entire sequences can be used directly to synthesize DNA probes by methods known to the skilled artisan such as random primer DNA labeling, nick translation, or end-labeling techniques, or RNA probes using available *in vitro* transcription systems. In addition, specific primers can be designed and used to amplify a part or all of the instant sequences. The resulting amplification products can be labeled directly during amplification reactions or labeled after amplification reactions, and used as probes to isolate full length cDNA or genomic fragments under conditions of appropriate stringency.

In addition, two short segments of the instant nucleic acid fragments may be used in polymerase chain reaction protocols to amplify longer nucleic acid fragments encoding

homologous genes from DNA or RNA. The polymerase chain reaction may also be performed on a library of cloned nucleic acid fragments wherein the sequence of one primer is derived from the instant nucleic acid fragments, and the sequence of the other primer takes advantage of the presence of the polyadenylic acid tracts to the 3' end of the mRNA precursor encoding plant genes. Alternatively, the second primer sequence may be based upon sequences derived from the cloning vector. For example, the skilled artisan can follow the RACE protocol (Frohman et al., (1988) *PNAS USA 85*:8998) to generate cDNAs by using PCR to amplify copies of the region between a single point in the transcript and the 3' or 5' end. Primers oriented in the 3' and 5' directions can be designed from the instant sequences. Using commercially available 3' RACE or 5' RACE systems (BRL), specific 3' or 5' cDNA fragments can be isolated (Ohara et al., (1989) *PNAS USA 86*:5673; Loh et al., (1989) *Science 243*:217). Products generated by the 3' and 5' RACE procedures can be combined to generate full-length cDNAs (Frohman, M. A. and Martin, G. R., (1989) *Techniques 1*:165).

Availability of the instant nucleotide and deduced amino acid sequences facilitates immunological screening of cDNA expression libraries. Synthetic peptides representing portions of the instant amino acid sequences may be synthesized. These peptides can be used to immunize animals to produce polyclonal or monoclonal antibodies with specificity for peptides or proteins comprising the amino acid sequences. These antibodies can be then be used to screen cDNA expression libraries to isolate full-length cDNA clones of interest (Lerner, R. A. (1984) *Adv. Immunol. 36*:1; Maniatis).

The nucleic acid fragments of the instant invention may be used to create transgenic plants in which the disclosed anthranilate synthase alpha subunit, anthranilate synthase beta subunit or tryptophan synthase alpha subunit are present at higher or lower levels than normal or in cell types or developmental stages in which they are not normally found. This would have the effect of altering the level of tryptophan in those cells. Manipulation of the levels of some of the antranilate synthase alpha subunits will also result in changes in the response to pathogen attack. Because this pathway is not followed for the production of tryptophan in higher animals, these enzymes are very good candidates for the discovery of herbicides and fungicides.

Overexpression of the anthranilate synthase alpha subunit, the anthranilate synthase beta subunit or the tryptophan synthase alpha subunit proteins of the instant invention may be accomplished by first constructing a chimeric gene in which the coding region is operably linked to a promoter capable of directing expression of a gene in the desired tissues at the desired stage of development. For reasons of convenience, the chimeric gene may comprise promoter sequences and translation leader sequences derived from the same genes. 3' Non-coding sequences encoding transcription termination signals may also be provided. The instant chimeric gene may also comprise one or more introns in order to facilitate gene expression.

Plasmid vectors comprising the instant chimeric gene can then constructed. The choice of plasmid vector is dependent upon the method that will be used to transform host plants. The skilled artisan is well aware of the genetic elements that must be present on the plasmid vector in order to successfully transform, select and propagate host cells containing the chimeric gene. The skilled artisan will also recognize that different independent transformation events will result in different levels and patterns of expression (Jones et al., (1985) *EMBO J. 4*:2411-2418; De Almeida et al., (1989) *Mol. Gen. Genetics 218*:78-86), and thus that multiple events must be screened in order to obtain lines displaying the desired expression level and pattern. Such screening may be accomplished by Southern analysis of DNA, Northern analysis of mRNA expression, Western analysis of protein expression, or phenotypic analysis.

For some applications it may be useful to direct the instant tryptophan biosynthetic enzyme to different cellular compartments, or to facilitate its secretion from the cell. It is thus envisioned that the chimeric gene described above may be further supplemented by altering the coding sequence to encode anthranilate synthase alpha subunit, anthranilate synthase beta subunit or tryptophan synthase alpha subunit with appropriate intracellular targeting sequences such as transit sequences (Keegstra, K. (1989) *Cell 56*:247-253), signal sequences or sequences encoding endoplasmic reticulum localization (Chrispeels, J.J., (1991) *Ann. Rev. Plant Phys. Plant Mol. Biol. 42*:21-53), or nuclear localization signals (Raikhel, N. (1992) *Plant Phys.100*:1627-1632) added and/or with targeting sequences that are already present removed. While the references cited give examples of each of these, the list is not exhaustive and more targeting signals of utility may be discovered in the future.

It may also be desirable to reduce or eliminate expression of genes encoding anthranilate synthase alpha subunit, anthranilate synthase beta subunit or tryptophan synthase alpha subunit in plants for some applications. In order to accomplish this, a chimeric gene designed for co-suppression of the instant tryptophan biosynthetic enzyme can be constructed by linking a gene or gene fragment encoding an anthranilate synthase alpha subunit, anthranilate synthase beta subunit or tryptophan synthase alpha subunit to plant promoter sequences. Alternatively, a chimeric gene designed to express antisense RNA for all or part of the instant nucleic acid fragment can be constructed by linking the gene or gene fragment in reverse orientation to plant promoter sequences. Either the co-suppression or antisense chimeric genes could be introduced into plants via transformation wherein expression of the corresponding endogenous genes are reduced or eliminated.

The instant anthranilate synthase alpha subunit, anthranilate synthase beta subunit or tryptophan synthase alpha subunit (or portions thereof) may be produced in heterologous host cells, particularly in the cells of microbial hosts, and can be used to prepare antibodies to the these proteins by methods well known to those skilled in the art. The antibodies are useful for detecting anthranilate synthase alpha subunit, anthranilate synthase beta subunit or tryptophan synthase alpha subunit *in situ* in cells or *in vitro* in cell extracts. Preferred

heterologous host cells for production of the instant anthranilate synthase alpha subunit, anthranilate synthase beta subunit or tryptophan synthase alpha subunit are microbial hosts. Microbial expression systems and expression vectors containing regulatory sequences that direct high level expression of foreign proteins are well known to those skilled in the art. Any of these could be used to construct a chimeric gene for production of the instant anthranilate synthase alpha subunit, anthranilate synthase beta subunit or tryptophan synthase alpha subunit. This chimeric gene could then be introduced into appropriate microorganisms via transformation to provide high level expression of the encoded tryptophan biosynthetic enzyme. An example of a vector for high level expression of the instant anthranilate synthase alpha subunit, anthranilate synthase beta subunit or tryptophan synthase alpha subunit in a bacterial host is provided (Example 7).

Additionally, the instant anthranilate synthase alpha subunit, anthranilate synthase beta subunit or tryptophan synthase alpha subunit can be used as targets to facilitate design and/or identification of inhibitors of those enzymes that may be useful as herbicides. This is desirable because the anthranilate synthase alpha subunit, the anthranilate synthase beta subunit and the tryptophan synthase alpha subunit described herein catalyze various steps in tryptophan biosynthesis. Accordingly, inhibition of the activity of one or more of the enzymes described herein could lead to inhibition plant growth. Thus, the instant anthranilate synthase alpha subunit, anthranilate synthase beta subunit, tryptophan synthase alpha subunit could be appropriate for new herbicide discovery and design.

All or a substantial portion of the nucleic acid fragments of the instant invention may also be used as probes for genetically and physically mapping the genes that they are a part of, and as markers for traits linked to those genes. Such information may be useful in plant breeding in order to develop lines with desired phenotypes. For example, the instant nucleic acid fragments may be used as restriction fragment length polymorphism (RFLP) markers. Southern blots (Maniatis) of restriction-digested plant genomic DNA may be probed with the nucleic acid fragments of the instant invention. The resulting banding patterns may then be subjected to genetic analyses using computer programs such as MapMaker (Lander et at., (1987) *Genomics 1*:174-181) in order to construct a genetic map. In addition, the nucleic acid fragments of the instant invention may be used to probe Southern blots containing restriction endonuclease-treated genomic DNAs of a set of individuals representing parent and progeny of a defined genetic cross. Segregation of the DNA polymorphisms is noted and used to calculate the position of the instant nucleic acid sequence in the genetic map previously obtained using this population (Botstein, D. et al., (1980) *Am. J. Hum. Genet. 32*:314-331).

The production and use of plant gene-derived probes for use in genetic mapping is described in R. Bernatzky, R. and Tanksley, S. D. (1986) *Plant Mol. Biol. Reporter 4(1)*:37-41. Numerous publications describe genetic mapping of specific cDNA clones using the methodology outlined above or variations thereof. For example, F2 intercross

populations, backcross populations, randomly mated populations, near isogenic lines, and other sets of individuals may be used for mapping. Such methodologies are well known to those skilled in the art.

Nucleic acid probes derived from the instant nucleic acid sequences may also be used for physical mapping (i.e., placement of sequences on physical maps; *see* Hoheisel, J. D., et al., In: *Nonmammalian Genomic Analysis: A Practical Guide*, Academic press 1996, pp. 319-346, and references cited therein).

In another embodiment, nucleic acid probes derived from the instant nucleic acid sequences may be used in direct fluorescence *in situ* hybridization (FISH) mapping (Trask, B. J. (1991) *Trends Genet. 7*:149-154). Although current methods of FISH mapping favor use of large clones (several to several hundred KB; see Laan, M. et al. (1995) *Genome Research 5*:13-20), improvements in sensitivity may allow performance of FISH mapping using shorter probes.

A variety of nucleic acid amplification-based methods of genetic and physical mapping may be carried out using the instant nucleic acid sequences. Examples include allele-specific amplification (Kazazian, H. H. (1989) *J. Lab. Clin. Med. 114(2)*:95-96), polymorphism of PCR-amplified fragments (CAPS; Sheffield, V. C. et al. (1993) *Genomics 16*:325-332), allele-specific ligation (Landegren, U. et al. (1988) *Science 241*:1077-1080), nucleotide extension reactions (Sokolov, B. P. (1990) *Nucleic Acid Res. 18*:3671), Radiation Hybrid Mapping (Walter, M. A. et al. (1997) *Nature Genetics 7*:22-28) and Happy Mapping (Dear, P. H. and Cook, P. R. (1989) *Nucleic Acid Res. 17*:6795-6807). For these methods, the sequence of a nucleic acid fragment is used to design and produce primer pairs for use in the amplification reaction or in primer extension reactions. The design of such primers is well known to those skilled in the art. In methods employing PCR-based genetic mapping, it may be necessary to identify DNA sequence differences between the parents of the mapping cross in the region corresponding to the instant nucleic acid sequence. This, however, is generally not necessary for mapping methods.

Loss of function mutant phenotypes may be identified for the instant cDNA clones either by targeted gene disruption protocols or by identifying specific mutants for these genes contained in a maize population carrying mutations in all possible genes (Ballinger and Benzer, (1989) *Proc. Natl. Acad. Sci USA 86*:9402; Koes et al., (1995) *Proc. Natl. Acad. Sci USA 92*:8149; Bensen et al., (1995) *Plant Cell 7*:75). The latter approach may be accomplished in two ways. First, short segments of the instant nucleic acid fragments may be used in polymerase chain reaction protocols in conjunction with a mutation tag sequence primer on DNAs prepared from a population of plants in which Mutator transposons or some other mutation-causing DNA element has been introduced (see Bensen, *supra*). The amplification of a specific DNA fragment with these primers indicates the insertion of the mutation tag element in or near the plant gene encoding the anthranilate synthase alpha subunit, the anthranilate synthase beta subunit or the tryptophan synthase alpha subunit.

Alternatively, the instant nucleic acid fragment may be used as a hybridization probe against PCR amplification products generated from the mutation population using the mutation tag sequence primer in conjunction with an arbitrary genomic site primer, such as that for a restriction enzyme site-anchored synthetic adaptor. With either method, a plant containing a mutation in the endogenous gene encoding an anthranilate synthase alpha subunit, an anthranilate synthase beta subunit or a tryptophan synthase alpha subunit can be identified and obtained. This mutant plant can then be used to determine or confirm the natural function of the anthranilate synthase alpha subunit, the anthranilate synthase beta subunit or the tryptophan synthase alpha subunit gene product.

EXAMPLES

The present invention is further defined in the following Examples, in which all parts and percentages are by weight and degrees are Celsius, unless otherwise stated. It should be understood that these Examples, while indicating preferred embodiments of the invention, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions.

EXAMPLE 1

Composition of cDNA Libraries; Isolation and Sequencing of cDNA Clones

cDNA libraries representing mRNAs from various corn, rice, soybean and wheat tissues were prepared. The characteristics of the libraries are described below.

TABLE 2

cDNA Libraries from Corn, Rice, Soybean and Wheat

Library	Tissue	Clone
cbn2	Corn Developing Kernel Two Days After Pollination	cbn2.pk0049.d2
cco1n	Corn Cob of 67 Day Old Plants Grown in Green House*	cco1n.pk0038.b1
cde1c	Corn Developing Embryo 20 Days After Pollination	cde1c.pk004.g2
ceb3	Corn Embryo 20 Days After Pollination	ceb3.pk0011.c12
cen3n	Corn Endosperm 20 Days After Pollination*	cen3n.pk0047.g5 cen3n.pk0210.c3 cen3n.pk0212.h6
cepe7	Corn 7 Day Old Epicotyl From Etiolated Seedling	cepe7.pk0012.g2
chp2	Corn (B73 and MK593) 11 Day Old Leaf Treated 24 Hours With Herbicides**	chp2.pk0020.e2
cpl1c	Corn Pooled BMS Treated With Chemical Chelators***	cpl1c.pk008.m4
cr1n	Corn Root From 7 Day Old Seedlings*	cr1n.pk0052.b8

Library	Tissue	Clone
csi1n	Corn Silk*	csi1n.pk0017.d4
cta1n	Corn Tassel*	cta1n.pk0051.g3
ctn1c	Corn Tassel, Night Harvested	ctn1c.pk001.l10
lkr	Corn 19 Day Old Seed	lkr.pk0013.g1
p0005	Corn Immature Ear	p0005.cbmew77r
p0015	Corn Embryo 13 Days After Pollination	p0015.cdpfk12r
p0031	Corn (CM45) Shoot Culture	p0031.ccmai08r p0031.ccmar07r
p0032	Corn Regenerating Callus (Hi-II 223a and 1129e), 10 and 14 Days After Auxin Removal, Pooled	p0032.crcas26r
p0096	Corn Scutelum 2 and 3 Days After Germinating	p0096.cnamt32r
p0102	Corn Early Meiosis Tassels*	p0102.ceraq01r
p0115	Corn Leaf and Sheath Meristem Tissue Collected from10th,11th, and 12th Leaves, Pooled	p0115.clsmj70r
p0125	Corn Anther Prophase I	p0125.czabf83r
p0126	Corn Leaf Tissue (V8-V10****), Night-Harvested	p0126.cnlcu75r
p0128	Corn Primary and Secondary Immature Ear	p0128.cpibl07r
rca1n	Rice Callus*	rca1n.pk004.p22
rds2c	Rice Developing Seeds	rds2c.pk004.i9
res1c	Rice Etiolated Seedling	res1c.pk004.c4
rl0n	Rice 15 Day Old Leaf*	rl0n.pk113.b4
rls24	Rice Leaf 15 Days After Germination, 24 Hours After Infection of Strain *Magaporthe grisea* 4360-R-67 (AVR2-YAMO); Susceptible	rls24.pk0002.e2
rr1	Rice Root of Two Week Old Developing Seedling	rr1.pk0038.h6 rr1.pk0071.f2
sdp3c	Soybean Developing Pods (8-9 mm)	sdp3c.pk019.c6
sfl1	Soybean Immature Flower	sfl1.pk129.d10
sls2c	Soybean Infected With *Sclerotinia sclerotiorum* mycelium	sls2c.pk025.d11
src3c	Soybean 8 Day Old Root Infected With Cyst Nematode	src3c.pk004.c19
wl1n	Wheat Leaf From 7 Day Old Seedling*	wl1n.pk0109.d5 wl1n.pk0110.c1
wle1n	Wheat Leaf From 7 Day Old Etiolated Seedling*	wle1n.pk0075.b4
wlm0	Wheat Seedlings 0 Hour After Inoculation With *Erysiphe graminis f. sp tritici*	wlm0.pk0028.f4
wlm96	Wheat Seedlings 96 Hours After Inoculation With *Erysiphe graminis f. sp tritici*	wlm96.pk026.j1 wlm96.pk041.h3 wlm96.pk045.c11

*These libraries were normalized essentially as described in U.S. Pat. No. 5,482,845

**Application of 2-[(2,4-dihydro-2,6,9-trimethyl[1]benzothiopyrano[4,3-*c*]pyrazol-8-yl)carbonyl]-1,3-cyclohexanedione *S*,*S*-dioxide (synthesis and methods of using this compound are described in WO 97/19087, incorporated herein by reference) and 2-[(2,3-dihydro-5,8-dimethylspiro[4*H*-1-benzothiopyran-4,2'-[1,3]dioxolan]-6-yl)carbonyl]-1,3-cyclohexanedione *S*,*S*-dioxide (also named 2-[(2,3-dihydro-5,8-dimethylspiro[4*H*-1-benzothiopyran-4,2'-[1,3]dioxolan]-6-yl)carbonyl]-3-hydroxy-2-cyclohexen-1-one *S*,*S*-dioxide; synthesis and methods of using this compound are described in WO 97/01550, incorporated herein by reference).

***Chemicals used included nitrilotriacetic acid, mercaptobenzothiazole, diethyldithiocarbamate

****For description of corn stages see Iowa State University Cooperative Extension Service Special Report No. 48.

cDNA libraries were prepared in Uni-ZAP™ XR vectors according to the manufacturer's protocol (Stratagene Cloning Systems, La Jolla, CA). Conversion of the Uni-ZAP™ XR libraries into plasmid libraries was accomplished according to the protocol provided by Stratagene. Upon conversion, cDNA inserts were contained in the plasmid vector pBluescript. cDNA inserts from randomly picked bacterial colonies containing recombinant pBluescript plasmids were amplified via polymerase chain reaction using primers specific for vector sequences flanking the inserted cDNA sequences or plasmid DNA was prepared from cultured bacterial cells. Amplified insert DNAs or plasmid DNAs were sequenced in dye-primer sequencing reactions to generate partial cDNA sequences (expressed sequence tags or "ESTs"; see Adams, M. D. et al., (1991) *Science 252*:1651). The resulting ESTs were analyzed using a Perkin Elmer Model 377 fluorescent sequencer.

EXAMPLE 2

Identification of cDNA Clones

ESTs encoding tryptophan biosynthetic enzymes were identified by conducting BLAST (Basic Local Alignment Search Tool; Altschul, S. F., et al., (1993) *J. Mol. Biol. 215*:403-410; see also www.ncbi.nlm.nih.gov/BLAST/) searches for similarity to sequences contained in the BLAST "nr" database (comprising all non-redundant GenBank CDS translations, sequences derived from the 3-dimensional structure Brookhaven Protein Data Bank, the last major release of the SWISS-PROT protein sequence database, EMBL, and DDBJ databases). The cDNA sequences obtained in Example 1 were analyzed for similarity to all publicly available DNA sequences contained in the "nr" database using the BLASTN algorithm provided by the National Center for Biotechnology Information (NCBI). The DNA sequences were translated in all reading frames and compared for similarity to all publicly available protein sequences contained in the "nr" database using the BLASTX algorithm (Gish, W. and States, D. J. (1993) *Nature Genetics 3*:266-272) provided by the NCBI. For convenience, the P-value (probability) of observing a match of a cDNA sequence to a sequence contained in the searched databases merely by chance as calculated by BLAST are reported herein as "pLog" values, which represent the negative of the

logarithm of the reported P-value. Accordingly, the greater the pLog value, the greater the likelihood that the cDNA sequence and the BLAST "hit" represent homologous proteins.

EXAMPLE 3

Characterization of cDNA Clones Encoding Anthranilate Synthase

The BLASTX search using the EST sequences from several corn, rice and wheat clones revealed similarity of the proteins encoded by the cDNAs to anthranilate synthase from *Arabidopsis thaliana* and *Ruta graveolens*. The BLASTX search using the nucleotide sequence from clone wre1n.pk0036.g10 also revealed similarity of the protein encoded by the cDNA to para-aminobenzoate synthase from *Streptomyces pristinaespiralis* (GenBank Accession No.U60417; pLog = 27.39). In the process of comparing the corn ESTs it was found that clones cen3n.pk0210.c3, cta1n.pk0025.f8 and cen3n.pk0047.g5 had overlapping regions of homology. A comparison of the corn ESTs from clones cen3n.pk0212.h6 and csi1n.pk0017.d4 also had overlapping regions of homology. Using this homology it was possible to align the ESTs and assemble two contigs (a contig is an assemblage of overlapping nucleic acid sequences to form one contiguous nucleotide sequence). The individual sequences were assembled into unique contiguous nucleotide sequences encoding anthranilate synthase from corn. The database accession numbers and BLAST results for each of these ESTs and contigs are shown in Table 3:

TABLE 3

BLAST Results for Clones Encoding Polypeptides Homologous to Anthranilate Synthase

Clone	Organism	Database Accession No	Blast Score pLog
Contig of: cen3n.pk0210.c3b cta1n.pk0025.f8 cen3n.pk0047.g5b	*A. thaliana*	GenBank L22585	70.69
ceb3.pk0011.c12	*R. graveolens*	GenBank L34344	48.15
cta1n.pk0051.g3	*R. graveolens*	GenBank L34344	32.22
cbn2.pk0049.d2	*R. graveolens*	GenBank L34344	35.00
Contig of: cen3n.pk0212.h6 csi1n.pk0017.d4	*A. thaliana*	GenBank L22585	34.00
cco1n.pk0038.b1	*A. thaliana*	GenBank L22585	17.69
rls24.pk0002.e2	*A. thaliana*	GenBank L22585	20.69
wle1n.pk0075.b4	*R. graveolens*	GenBank L34343	38.00

Characterization of cDNA Clones Encoding Anthranilate Synthase Alpha Subunit

The sequence of the entire cDNA insert in clone cta1n.pk0051.g3 was determined. A contig was assembled with this sequence and sequences from portions of the cDNA inserts

in clones cde1c.pk004.g2, p0125.czabf83r, p0005.cbmew77r, p0096.cnamt32r, p0032.crcas26r, p0031.ccmar07r, p0015.cdpfk12r and p0128.cpibl07r. The sequence of the entire cDNA insert in clone cta1n.pk0051.g3 was determined. A contig was assembled with this sequence and sequences from portions of the cDNA in clones p0115.clsmj70r, cbn2.pk0049.d2, cpl1c.pk008.m4 and p0102.ceraq01r. The BLASTX search using these contig sequences and the EST sequences from clones rca1n.pk004.p22, sdp3c.pk019.c6, sfl1.pk129.d10, wle1n.pk0075.b4, the 5' sequence from clone wlm96.pk026.j1 and the 3' sequence from clone wlm96.pk026.j1 revealed similarity of the protein encoded by the cDNA to anthranilate synthase alpha subunit from *Ruta graveolens* (NCBI General Identifier No. 960289). The BLAST results for each of these sequences are shown in Table 4:

TABLE 4

BLAST Results for Clones Encoding Polypeptides Homologous to Anthranilate Synthase Alpha Subunit

Clone	BLAST pLog Score NCBI General Identifier No. 960289
Contig of: cde1c.pk004.g2 p0125.czabf83r p0005.cbmew77r p0096.cnamt32r p0032.crcas26r p0031.ccmar07r p0015.cdpfk12r p0128.cpibl07r cta1n.pk0051.g3	>254
Contig of: p0115.clsmj70r cbn2.pk0049.d2 cpl1c.pk008.m4 p0102.ceraq01r ceb3.pk0011.c12	>254
rca1n.pk004.p22	20.15
sdp3c.pk019.c6	48.52
sfl1.pk129.d10	34.70
wle1n.pk0075.b4 (5'end)	57.00
wle1n.pk0075.b4 (3'end)	35.70
wlm96.pk026.j1	5.40

The sequence of the contig assembled from the entire cDNA insert from clone cta1n.pk0051.g3 and a portion of the cDNA insert in clones cde1c.pk004.g2, p0125.czabf83r, p0005.cbmew77r, p0096.cnamt32r, p0032.crcas26r, p0031.ccmar07r, p0015.cdpfk12r and p0128.cpibl07r is shown in SEQ ID NO:1; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:2. The amino acid sequence set forth in

SEQ ID NO:2 was evaluated by BLASTP, yielding a pLog value >254 versus the *Ruta graveolens* sequence. The sequence of the contig assembled from the entire cDNA insert in clone cta1n.pk0051.g3 and a portion of the cDNA in clones p0115.clsmj70r, cbn2.pk0049.d2, cpl1c.pk008.m4 and p0102.ceraq01r is shown in SEQ ID NO:3; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:4. The sequence of a portion of the cDNA insert from clone rca1n.pk004.p22 is shown in SEQ ID NO:5; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:6. The sequence of a portion of the cDNA insert from clone sdp3c.pk019.c6 is shown in SEQ ID NO:7; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:8. The sequence of a portion of the cDNA insert from clone sfl1.pk129.d10 is shown in SEQ ID NO:9; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:10. The sequence of the 5' terminal portion of the cDNA insert from clone wle1n.pk0075.b4 is shown in SEQ ID NO:11; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:12. The sequence of the 3' terminal portion of the cDNA insert from clone wle1n.pk0075.b4 is shown in SEQ ID NO:13; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:14. The sequence of a portion of the cDNA insert from clone wlm96.pk026.j1 is shown in SEQ ID NO:15; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:16.

Figure 2 presents an alignment of the amino acid sequence set forth in SEQ ID NO:2 and the *Ruta graveolens* anthranilate synthase alpha subunit sequence (SEQ ID NO:39). The amino acid sequence set forth in SEQ ID NO:2 is 64.7% similar to the *Ruta graveolens* sequence. Sequence alignments and percent similarity calculations were performed using the Megalign program of the LASARGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Pairwise alignment of the sequences was performed using the Clustal method of alignment (Higgins, D.G. and Sharp, P.M. (1989) *CABIOS. 5*:151-153) with the default parameters (KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5).

Sequence alignments and BLAST scores and probabilities indicate that the instant nucleic acid fragments encode an entire corn anthranilate synthase alpha subunit and portions of corn, rice, soybean and wheat anthranilate synthase alpha subunits. These sequences represent the first corn, rice, soybean and wheat sequences encoding anthranilate synthase alpha subunit.

Characterization of cDNA Clones Encoding Anthranilate Synthase Beta Subunit

The sequence of a larger portion of the cDNA insert in clone cen3n.pk0210.c3 was determined, this sequence includes the sequence from clone cta1n.pk0025.f8. A contig was assembled with the sequence from clone cen3n.pk0210.c3 and sequence from a portion of the cDNA in clones cepe7.pk0012.g2 and cen3n.pk0047.g5. The BLASTX search using these contig sequences, the sequences from the contig assembled of clones p0126.cnlcu75r, cen3n.pk0212.h6, cco1n.pk0038.b1, csi1n.pk0017.d4, p0031.ccmai08r and ctn1c.pk001.l10,

the contig assembled from clones rds2c.pk004.i9, rls24.pk0002.e2 and rl0n.pk113.b4 and the EST sequences from clone wlm96.pk045.c11 revealed similarity of the proteins encoded by the cDNAs to anthranilate synthase beta subunit from *Arabidopsis thaliana* (NCBI General Identifier No. 541849). The BLAST results for each of these ESTs are shown in Table 5:

TABLE 5

BLAST Results for Clones Encoding Polypeptides Homologous to Anthranilate Synthase Beta Subunit

Clone	BLAST pLog Score NCBI General Identifier No. 541849
Contig of: cen3n.pk0210.c3 cepe7.pk0012.g2 cen3n.pk0047.g5	84.52
Contig of: p0126.cnlcu75r cen3n.pk0212.h6 cco1n.pk0038.b1 csi1n.pk0017.d4 p0031.ccmai08r ctn1c.pk001.l10	110.0
Contig of: rds2c.pk004.i9 rls24.pk0002.e2 rl0n.pk113.b4	20.15
wlm96.pk045.c11	15.00

The sequence of the contig assembled from a portion of the cDNA insert from clones cen3n.pk0210.c3, cepe7.pk0012.g2 and cen3n.pk0047.g5 is shown in SEQ ID NO:17; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:18. The sequence of the contig assembled from a portion of the cDNA insert from clones p0126.cnlcu75r, cen3n.pk0212.h6, cco1n.pk0038.b1, csi1n.pk0017.d4, p0031.ccmai08r and ctn1c.pk001.l10 is shown in SEQ ID NO:17; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:18. The amino acid sequence set forth in SEQ ID NO:18 was evaluated by BLASTP, yielding a pLog value of 92.22 versus the *Arabidopsis thaliana* sequence. The sequence of the contig assembled from a portion of the cDNA insert from clones rds2c.pk004.i9, rls24.pk0002.e2 and rl0n.pk113.b4 is shown in SEQ ID NO:19; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:20. The sequence of a portion of the cDNA insert from clone wlm96.pk045.c11 is shown in SEQ ID NO:21; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:22.

Figure 3 presents an alignment of the amino acid sequences set forth in SEQ ID NOs:20 and 22 and the *Arabidopsis thaliana* anthranilate synthase beta subunit sequence (SEQ ID NO:40). The amino acid sequences set forth in SEQ ID NO:20 are 58.0% similar to the *Arabidopsis thaliana* sequence while the amino acid sequences set forth in SEQ ID

NO:22 are 57.1% similar to the *Arabidopsis thaliana* sequence. Sequence alignments and percent similarity calculations were performed using the Megalign program of the LASARGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins, D.G. and Sharp, P.M. (1989) *CABIOS. 5*:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10).

Sequence alignments and BLAST scores and probabilities indicate that the instant nucleic acid fragments encode an entire corn anthranilate synthase beta subunit, an almost entire rice anthranilate synthase beta subunit and portions of corn and wheat anthranilate synthase beta subunit. These sequences represent the first corn, rice and wheat sequences encoding anthranilate synthase beta subunit.

EXAMPLE 4

Characterization of cDNA Clones Encoding Tryptophan Synthase

The BLASTX search using the EST sequences from several corn, rice, soybean and wheat clones revealed similarity of the proteins encoded by the cDNAs to tryptophan synthase from *Zea mays* and *Arabidopsis thaliana*. In the process of comparing the rice ESTs it was found that clones wl1n.pk0109.d5 and wl1n.pk0110.c1 had overlapping regions of homology. A comparison of the corn ESTs from clones chp2.pk0020.e2, cen3n.pk0147.d4, lkr.pk0013.g1, cs1/pk0089.a8, cta1n.pk0057.d6, cen3n.pk0014.c5, cen3n.pk0193.c5 and cen3n.pk0135.d1 also revealed overlapping regions of homology. A comparison of the corn ESTs from clones m.15.6.g07.sk20, m.15.4.g07.sk20 and cco1n.pk0030.g7 revealed overlapping regions of homology. Using this homology it was possible to align the ESTs and assemble four contigs (a contig is an assemblage of overlapping nucleic acid sequences to form one contiguous nucleotide sequence). The individual sequences were assembled into unique contiguous nucleotide sequences encoding tryptophan synthase from corn, rice, soybean and wheat. The database accession numbers and BLAST results for each of these ESTs and contigs are shown in Table 6:

TABLE 6

BLASTX Results for Clones Encoding Polypeptides Homologous to Trypotophan Synthase

Clone	Organism	Database Accession No	Blast Score pLog
rr1.pk0038.h6	*A. thaliana*	GenBank U18993	26.00
Contig of: wl1n.pk0109.d5 wl1n.pk0110.c1	*Z. mays*	GenBank X76713	52.39
wl1n.pk0067.f2	*A. thaliana*	GenBank U18993	19.69
cr1n.pk0033.f8	*Z. mays*	GenBank X76713	47.69
cr1n.pk0052.b8	*Z. mays*	GenBank X76713	25.70

Clone	Organism	Database Accession No	Blast Score pLog
Contig of: chp2.pk0020.e2 cen3n.pk0147.d4 lkr.pk0013.g1 cs1.pk0089.a8 cta1n.pk0057.d6 cen3n.pk0014.c5 cen3n.pk0193.c5 cen3n.pk0135.d1	*Z mays*	GenBank X76713	83.40
Contig of: m.15.6.g07.sk20 m15.4.g07.sk20 cco1n.pk0030.g7	*Z. mays*	GenBank X76713	25.52

Characterization of cDNA Clones Encoding Tryptophan Synthase Alpha Subunit

The sequence of the entire cDNA insert in clone lkr.pk0013.g1 and a larger portion of the sequence from clone chp2.pk0020.e2 have been determined. A contig has been assembled with these two sequences and they have been found to include the sequences from clones cco1n.pk0030.g7, m.15.6.g07.sk20, cen3n.pk0014.c5, cen3n.pk0135.d1, cen3n.pk0147.d4, cen3n.pk0193.c5, cta1n.pk0057.d6, cs1.pk0089.a8 and m.15.6.g07.sk20. The sequence from the entire cDNA insert in clone cr1n.pk0052.b8 has been determined, it includes the sequences from clone cr1n.pk0033.f8. The BLASTX search using these sequences, the sequence from the contig assembled of clones res1c.pk004.c4 and rr1.pk0071.f2, the EST sequences from clone rr1.pk0038.h6 and the sequence of the contig assembled from clones src3c.pk004.c19 and sls2c.pk025.d11 revealed similarity of the proteins encoded by the cDNAs to tryptophan synthase alpha subunit from *Arabidopsis thaliana* (NCBI General Identifier No. 2129755). The BLASTX search using the sequence of the contig assembled from clones wl1n.pk0109.d5 and wl1n.pk0110.c1 revealed similarity of the proteins encoded by the cDNAs to tryptophan synthase alpha subunit from *Zea mays* (NCBI General Identifier No. 1174783). The BLAST results for each of these sequences are shown in Table 7:

TABLE 7

BLAST Results for Clones Encoding Polypeptides Homologous to Tryptophan Synthase Alpha Subunit

Clone	NCBI General Identifier No.	BLAST pLog Score
Contig of: chp2.pk0020.e2 lkr.pk0013.g1	2129755	124.0
cr1n.pk0052.b8	2129755	120.0
Contig of: res1c.pk004.c4 rr1.pk0071.f2	2129755	78.70
rr1.pk0038.h6	2129755	125.0
Contig of: src3c.pk004.c19 sls2c.pk025.d11	2129755	49.00
Contig of: wlm0.pk0028.f4 wlm96.pk041.h3	1174783	33.70
Contig of: wl1n.pk0109.d5 wl1n.pk0110.c1	1174783	85.70

The sequence of the contig assembled from the entire cDNA insert in clone lkr.pk0013.g1 and a large portion of the cDNA insert in clone chp2.pk0020.e2 is shown in SEQ ID NO:25; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:26. The amino acid sequence set forth in SEQ ID NO:26 was evaluated by BLASTP, yielding a pLog value of 105.0 versus the *Arabidopsis thaliana* sequence. The sequence of the entire cDNA insert in clone cr1n.pk0052.b8 was determined and is shown in SEQ ID NO:27; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:28. The amino acid sequence set forth in SEQ ID NO:28 was evaluated by BLASTP, yielding a pLog value of 98.22 versus the *Arabidopsis thaliana* sequence. The sequence of the contig assembled from a portion of the cDNA insert in clones res1c.pk004.c4 and rr1.pk0071.f2 is shown in SEQ ID NO:29; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:30. The sequence of the entire cDNA insert in clone rr1.pk0038.h6 was determined and is shown in SEQ ID NO:31; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:32. The amino acid sequence set forth in SEQ ID NO:32 was evaluated by BLASTP, yielding a pLog value of 106.0 versus the *Arabidopsis thaliana* sequence. The sequence of the contig assembled from a portion of the cDNA insert from clones src3c.pk004.c19 and sls2c.pk025.d11 is shown in SEQ ID NO:33; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:34. The sequence of the contig assembled from a portion of the cDNA insert from clones wlm0.pk0028.f4 and wlm96.pk041.h3 is shown in SEQ ID NO:35; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:36. The

sequence of the contig assembled from a portion of the cDNA insert from clones wl1n.pk0109.d5 and wl1n.pk0110.c1 is shown in SEQ ID NO:37; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:38.

Figure 3 presents an alignment of the amino acid sequences set forth in SEQ ID NOs:26, 28 and 32 and the *Arabidopsis thaliana* sequence (SEQ ID NO:41). The data in Table 8 represents a calculation of the percent identity of the amino acid sequences set forth in SEQ ID NOs:26, 28 and 32 and the *Arabidopsis thaliana* tryptophan synthase alpha subunit sequence.

TABLE 8

Percent Identity of Amino Acid Sequences Deduced From the Nucleotide Sequences of cDNA Clones Encoding Polypeptides Homologous to Tryptophan Synthase Alpha Subunit

Clone	SEQ ID NO.	Percent Identity to 2129755
Contig of: chp2.pk0020.e2 lkr.pk0013.g1	26	59.0
cr1n.pk0052.b8	28	56.7
rr1.pk0038.h6	32	59.9

Sequence alignments and percent similarity calculations were performed using the Megalign program of the LASARGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins, D.G. and Sharp, P.M. (1989) *CABIOS.* 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10).

Sequence alignments and BLAST scores and probabilities indicate that the instant nucleic acid fragments encode entire or nearly entire corn and rice tryptophan synthase alpha subunit and portions of rice, soybean and wheat tryptophan synthase alpha subunit. These sequences represent the first rice, soybean and wheat and variant corn sequences encoding tryptophan synthase alpha subunit.

EXAMPLE 5

Expression of Chimeric Genes in Monocot Cells

A chimeric gene comprising a cDNA encoding a tryptophan biosynthetic enzyme in sense orientation with respect to the maize 27 kD zein promoter that is located 5' to the cDNA fragment, and the 10 kD zein 3' end that is located 3' to the cDNA fragment, can be constructed. The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites (NcoI or SmaI) can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the digested vector pML103 as described below. Amplification is then performed in a standard PCR. The amplified DNA is then digested

with restriction enzymes NcoI and SmaI and fractionated on an agarose gel. The appropriate band can be isolated from the gel and combined with a 4.9 kb NcoI-SmaI fragment of the plasmid pML103. Plasmid pML103 has been deposited under the terms of the Budapest Treaty at ATCC (American Type Culture Collection, 10801 University Blvd., Manassas, VA 20110-2209), and bears accession number ATCC 97366. The DNA segment from pML103 contains a 1.05 kb SalI-NcoI promoter fragment of the maize 27 kD zein gene and a 0.96 kb SmaI-SalI fragment from the 3' end of the maize 10 kD zein gene in the vector pGem9Zf(+) (Promega). Vector and insert DNA can be ligated at 15°C overnight, essentially as described (Maniatis). The ligated DNA may then be used to transform *E. coli* XL1-Blue (Epicurian Coli XL-1 Blue™; Stratagene). Bacterial transformants can be screened by restriction enzyme digestion of plasmid DNA and limited nucleotide sequence analysis using the dideoxy chain termination method (Sequenase™ DNA Sequencing Kit; U. S. Biochemical). The resulting plasmid construct would comprise a chimeric gene encoding, in the 5' to 3' direction, the maize 27 kD zein promoter, a cDNA fragment encoding a tryptophan biosynthetic enzyme, and the 10 kD zein 3' region.

The chimeric gene described above can then be introduced into corn cells by the following procedure. Immature corn embryos can be dissected from developing caryopses derived from crosses of the inbred corn lines H99 and LH132. The embryos are isolated 10 to 11 days after pollination when they are 1.0 to 1.5 mm long. The embryos are then placed with the axis-side facing down and in contact with agarose-solidified N6 medium (Chu et al., (1975) *Sci. Sin. Peking* 18:659-668). The embryos are kept in the dark at 27°C. Friable embryogenic callus consisting of undifferentiated masses of cells with somatic proembryoids and embryoids borne on suspensor structures proliferates from the scutellum of these immature embryos. The embryogenic callus isolated from the primary explant can be cultured on N6 medium and sub-cultured on this medium every 2 to 3 weeks.

The plasmid, p35S/Ac (obtained from Dr. Peter Eckes, Hoechst Ag, Frankfurt, Germany) may be used in transformation experiments in order to provide for a selectable marker. This plasmid contains the *Pat* gene (see European Patent Publication 0 242 236) which encodes phosphinothricin acetyl transferase (PAT). The enzyme PAT confers resistance to herbicidal glutamine synthetase inhibitors such as phosphinothricin. The *pat* gene in p35S/Ac is under the control of the 35S promoter from Cauliflower Mosaic Virus (Odell et al. (1985) *Nature* 313:810-812) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*.

The particle bombardment method (Klein et al., (1987) *Nature* 327:70-73) may be used to transfer genes to the callus culture cells. According to this method, gold particles (1 μm in diameter) are coated with DNA using the following technique. Ten μg of plasmid DNAs are added to 50 μL of a suspension of gold particles (60 mg per mL). Calcium chloride (50 μL of a 2.5 M solution) and spermidine free base (20 μL of a 1.0 M solution) are added to the particles. The suspension is vortexed during the addition of these solutions.

After 10 minutes, the tubes are briefly centrifuged (5 sec at 15,000 rpm) and the supernatant removed. The particles are resuspended in 200 μL of absolute ethanol, centrifuged again and the supernatant removed. The ethanol rinse is performed again and the particles resuspended in a final volume of 30 μL of ethanol. An aliquot (5 μL) of the DNA-coated gold particles can be placed in the center of a Kapton™ flying disc (Bio-Rad Labs). The particles are then accelerated into the corn tissue with a Biolistic™ PDS-1000/He (Bio-Rad Instruments, Hercules CA), using a helium pressure of 1000 psi, a gap distance of 0.5 cm and a flying distance of 1.0 cm.

For bombardment, the embryogenic tissue is placed on filter paper over agarose-solidified N6 medium. The tissue is arranged as a thin lawn and covered a circular area of about 5 cm in diameter. The petri dish containing the tissue can be placed in the chamber of the PDS-1000/He approximately 8 cm from the stopping screen. The air in the chamber is then evacuated to a vacuum of 28 inches of Hg. The macrocarrier is accelerated with a helium shock wave using a rupture membrane that bursts when the He pressure in the shock tube reaches 1000 psi.

Seven days after bombardment the tissue can be transferred to N6 medium that contains gluphosinate (2 mg per liter) and lacks casein or proline. The tissue continues to grow slowly on this medium. After an additional 2 weeks the tissue can be transferred to fresh N6 medium containing gluphosinate. After 6 weeks, areas of about 1 cm in diameter of actively growing callus can be identified on some of the plates containing the glufosinate-supplemented medium. These calli may continue to grow when sub-cultured on the selective medium.

Plants can be regenerated from the transgenic callus by first transferring clusters of tissue to N6 medium supplemented with 0.2 mg per liter of 2,4-D. After two weeks the tissue can be transferred to regeneration medium (Fromm et al., (1990) *Bio/Technology 8*:833-839).

EXAMPLE 6

Expression of Chimeric Genes in Dicot Cells

A seed-specific expression cassette composed of the promoter and transcription terminator from the gene encoding the β subunit of the seed storage protein phaseolin from the bean *Phaseolus vulgaris* (Doyle et al. (1986) *J. Biol. Chem. 261*:9228-9238) can be used for expression of the instant tryptophan biosynthetic enzyme in transformed soybean. The phaseolin cassette includes about 500 nucleotides upstream (5') from the translation initiation codon and about 1650 nucleotides downstream (3') from the translation stop codon of phaseolin. Between the 5' and 3' regions are the unique restriction endonuclease sites Nco I (which includes the ATG translation initiation codon), Sma I, Kpn I and Xba I. The entire cassette is flanked by Hind III sites.

The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites can be

incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the expression vector. Amplification is then performed as described above, and the isolated fragment is inserted into a pUC18 vector carrying the seed expression cassette.

Soybean embroys may then be transformed with the expression vector comprising sequences encoding a tryptophan biosynthetic enzyme. To induce somatic embryos, cotyledons, 3-5 mm in length dissected from surface sterilized, immature seeds of the soybean cultivar A2872, can be cultured in the light or dark at 26°C on an appropriate agar medium for 6-10 weeks. Somatic embryos which produce secondary embryos are then excised and placed into a suitable liquid medium. After repeated selection for clusters of somatic embryos which multiplied as early, globular staged embryos, the suspensions are maintained as described below.

Soybean embryogenic suspension cultures can maintained in 35 mL liquid media on a rotary shaker, 150 rpm, at 26°C with florescent lights on a 16:8 hour day/night schedule. Cultures are subcultured every two weeks by inoculating approximately 35 mg of tissue into 35 mL of liquid medium.

Soybean embryogenic suspension cultures may then be transformed by the method of particle gun bombardment (Kline et al. (1987) *Nature* (London) *327*:70, U.S. Patent No. 4,945,050). A DuPont Biolistic™ PDS1000/HE instrument (helium retrofit) can be used for these transformations.

A selectable marker gene which can be used to facilitate soybean transformation is a chimeric gene composed of the 35S promoter from Cauliflower Mosaic Virus (Odell et al.(1985) *Nature 313*:810-812), the hygromycin phosphotransferase gene from plasmid pJR225 (from *E. coli*; Gritz et al.(1983) *Gene 25*:179-188) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*. The seed expression cassette comprising the phaseolin 5' region, the fragment encoding the tryptophan biosynthetic enzyme and the phaseolin 3' region can be isolated as a restriction fragment. This fragment can then be inserted into a unique restriction site of the vector carrying the marker gene.

To 50 µL of a 60 mg/mL 1 µm gold particle suspension is added (in order): 5 µL DNA (1 µg/µL), 20 µl spermidine (0.1 M), and 50 µL $CaCl_2$ (2.5 M). The particle preparation is then agitated for three minutes, spun in a microfuge for 10 seconds and the supernatant removed. The DNA-coated particles are then washed once in 400 µL 70% ethanol and resuspended in 40 µL of anhydrous ethanol. The DNA/particle suspension can be sonicated three times for one second each. Five µL of the DNA-coated gold particles are then loaded on each macro carrier disk.

Approximately 300-400 mg of a two-week-old suspension culture is placed in an empty 60x15 mm petri dish and the residual liquid removed from the tissue with a pipette. For each transformation experiment, approximately 5-10 plates of tissue are normally

bombarded. Membrane rupture pressure is set at 1100 psi and the chamber is evacuated to a vacuum of 28 inches mercury. The tissue is placed approximately 3.5 inches away from the retaining screen and bombarded three times. Following bombardment, the tissue can be divided in half and placed back into liquid and cultured as described above.

Five to seven days post bombardment, the liquid media may be exchanged with fresh media, and eleven to twelve days post bombardment with fresh media containing 50 mg/mL hygromycin. This selective media can be refreshed weekly. Seven to eight weeks post bombardment, green, transformed tissue may be observed growing from untransformed, necrotic embryogenic clusters. Isolated green tissue is removed and inoculated into individual flasks to generate new, clonally propagated, transformed embryogenic suspension cultures. Each new line may be treated as an independent transformation event. These suspensions can then be subcultured and maintained as clusters of immature embryos or regenerated into whole plants by maturation and germination of individual somatic embryos.

EXAMPLE 7

Expression of Chimeric Genes in Microbial Cells

The cDNAs encoding the instant tryptophan biosynthetic enzymes can be inserted into the T7 *E. coli* expression vector pBT430. This vector is a derivative of pET-3a (Rosenberg et al. (1987) *Gene 56*:125-135) which employs the bacteriophage T7 RNA polymerase/T7 promoter system. Plasmid pBT430 was constructed by first destroying the EcoR I and Hind III sites in pET-3a at their original positions. An oligonucleotide adaptor containing EcoR I and Hind III sites was inserted at the BamH I site of pET-3a. This created pET-3aM with additional unique cloning sites for insertion of genes into the expression vector. Then, the Nde I site at the position of translation initiation was converted to an Nco I site using oligonucleotide-directed mutagenesis. The DNA sequence of pET-3aM in this region, 5'-CATATGG, was converted to 5'-CCCATGG in pBT430.

Plasmid DNA containing a cDNA may be appropriately digested to release a nucleic acid fragment encoding the protein. This fragment may then be purified on a 1% NuSieve GTG™ low melting agarose gel (FMC). Buffer and agarose contain 10 μg/ml ethidium bromide for visualization of the DNA fragment. The fragment can then be purified from the agarose gel by digestion with GELase™ (Epicentre Technologies) according to the manufacturer's instructions, ethanol precipitated, dried and resuspended in 20 μL of water. Appropriate oligonucleotide adapters may be ligated to the fragment using T4 DNA ligase (New England Biolabs, Beverly, MA). The fragment containing the ligated adapters can be purified from the excess adapters using low melting agarose as described above. The vector pBT430 is digested, dephosphorylated with alkaline phosphatase (NEB) and deproteinized with phenol/chloroform as described above. The prepared vector pBT430 and fragment can then be ligated at 16°C for 15 hours followed by transformation into DH5 electrocompetent cells (GIBCO BRL). Transformants can be selected on agar plates containing LB media and 100 μg/mL ampicillin. Transformants containing the gene encoding the tryptophan

biosynthetic enzyme are then screened for the correct orientation with respect to the T7 promoter by restriction enzyme analysis.

For high level expression, a plasmid clone with the cDNA insert in the correct orientation relative to the T7 promoter can be transformed into *E. coli* strain BL21(DE3) (Studier et al. (1986) *J. Mol. Biol. 189*:113-130). Cultures are grown in LB medium containing ampicillin (100 mg/L) at 25°C. At an optical density at 600 nm of approximately 1, IPTG (isopropylthio-β-galactoside, the inducer) can be added to a final concentration of 0.4 mM and incubation can be continued for 3 h at 25°. Cells are then harvested by centrifugation and re-suspended in 50 µL of 50 mM Tris-HCl at pH 8.0 containing 0.1 mM DTT and 0.2 mM phenyl methylsulfonyl fluoride. A small amount of 1 mm glass beads can be added and the mixture sonicated 3 times for about 5 seconds each time with a microprobe sonicator. The mixture is centrifuged and the protein concentration of the supernatant determined. One µg of protein from the soluble fraction of the culture can be separated by SDS-polyacrylamide gel electrophoresis. Gels can be observed for protein bands migrating at the expected molecular weight.

EXAMPLE 8

Evaluating Compounds for Their Ability to Inhibit the Activity of Tryptophan Biosynthetic Enzymes

The tryptophan biosynthetic enzymes described herein may be produced using any number of methods known to those skilled in the art. Such methods include, but are not limited to, expression in bacteria as described in Example 7, or expression in eukaryotic cell culture, *in planta*, and using viral expression systems in suitably infected organisms or cell lines. The instant tryptophan biosynthetic enzymes may be expressed either as mature forms of the proteins as observed *in vivo* or as fusion proteins by covalent attachment to a variety of enzymes, proteins or affinity tags. Common fusion protein partners include glutathione S-transferase ("GST"), thioredoxin ("Trx"), maltose binding protein, and C- and/or N-terminal hexahistidine polypeptide ("$(His)_6$"). The fusion proteins may be engineered with a protease recognition site at the fusion point so that fusion partners can be separated by protease digestion to yield intact mature enzyme. Examples of such proteases include thrombin, enterokinase and factor Xa. However, any protease can be used which specifically cleaves the peptide connecting the fusion protein and the enzyme.

Purification of the instant tryptophan biosynthetic enzymes, if desired, may utilize any number of separation technologies familiar to those skilled in the art of protein purification. Examples of such methods include, but are not limited to, homogenization, filtration, centrifugation, heat denaturation, ammonium sulfate precipitation, desalting, pH precipitation, ion exchange chromatography, hydrophobic interaction chromatography and affinity chromatography, wherein the affinity ligand represents a substrate, substrate analog or inhibitor. When the tryptophan biosynthetic enzymes are expressed as fusion proteins, the purification protocol may include the use of an affinity resin which is specific for the fusion

protein tag attached to the expressed enzyme or an affinity resin containing ligands which are specific for the enzyme. For example, a tryptophan biosynthetic enzyme may be expressed as a fusion protein coupled to the C-terminus of thioredoxin. In addition, a $(His)_6$ peptide may be engineered into the N-terminus of the fused thioredoxin moiety to afford additional opportunities for affinity purification. Other suitable affinity resins could be synthesized by linking the appropriate ligands to any suitable resin such as Sepharose-4B. In an alternate embodiment, a thioredoxin fusion protein may be eluted using dithiothreitol; however, elution may be accomplished using other reagents which interact to displace the thioredoxin from the resin. These reagents include β–mercaptoethanol or other reduced thiol. The eluted fusion protein may be subjected to further purification by traditional means as stated above, if desired. Proteolytic cleavage of the thioredoxin fusion protein and the enzyme may be accomplished after the fusion protein is purified or while the protein is still bound to the ThioBond™ affinity resin or other resin.

Crude, partially purified or purified enzyme, either alone or as a fusion protein, may be utilized in assays for the evaluation of compounds for their ability to inhibit enzymatic activation of the tryptophan biosynthetic enzymes disclosed herein. Assays may be conducted under well known experimental conditions which permit optimal enzymatic activity. For example, assays for anthranilate synthase alpha subunit are presented by Bohlmann J et al. (1995) *Plant J* 7:491-501. Assays for anthranilate synthase beta subunit are presented by Walker M. S. and DeMoss J. A. (1983) *J Biol Chem 258*:3571-3575. Assays for tryptophan synthase alpha subunit are presented by Zhao J and Last R. L. (1995) *J Biol Chem 270*:6081-7.

CLAIMS

What is claimed is:

1. An isolated nucleic acid fragment encoding all or a substantial portion of an anthranilate synthase alpha subunit comprising a member selected from the group consisting of:

(a) an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQID NO:12, SEQ ID NO:14 and SEQ ID NO:16;

(b) an isolated nucleic acid fragment that is substantially similar to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQID NO:12, SEQ ID NO:14 and SEQ ID NO:16; and

(c) an isolated nucleic acid fragment that is complementary to (a) or (b).

2. The isolated nucleic acid fragment of Claim 1 wherein the isolated nucleic acid fragment hybridizes under stringent conditions to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in SEQ ID NO:2.

3. The isolated nucleic acid fragment of Claim 1 wherein the isolated nucleic acid fragment hybridizes under stringent conditions to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in SEQ ID NO:4.

4. The isolated nucleic acid fragment of Claim 1 wherein the isolated nucleic acid fragment hybridizes under stringent conditions to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in SEQ ID NO:6.

5. The isolated nucleic acid fragment of Claim 1 wherein the isolated nucleic acid fragment hybridizes under stringent conditions to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in SEQ ID NO:8.

6. The isolated nucleic acid fragment of Claim 1 wherein the isolated nucleic acid fragment hybridizes under stringent conditions to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in SEQ ID NO:10.

7. The isolated nucleic acid fragment of Claim 1 wherein the isolated nucleic acid fragment hybridizes under stringent conditions to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in SEQ ID NO:12.

8. The isolated nucleic acid fragment of Claim 1 wherein the isolated nucleic acid fragment hybridizes under stringent conditions to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in SEQ ID NO:14.

9. The isolated nucleic acid fragment of Claim 1 wherein the isolated nucleic acid fragment hybridizes under stringent conditions to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in SEQ ID NO:16.

10. The isolated nucleic acid fragment of Claim 1 wherein the isolated nucleic acid fragment is at least 80% similar to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in SEQ ID NO:2.

11. The isolated nucleic acid fragment of Claim 1 wherein the isolated nucleic acid fragment is at least 80% similar to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in SEQ ID NO:4.

12. The isolated nucleic acid fragment of Claim 1 wherein the isolated nucleic acid fragment is at least 80% similar to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in SEQ ID NO:6.

13. The isolated nucleic acid fragment of Claim 1 wherein the isolated nucleic acid fragment is at least 80% similar to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in SEQ ID NO:8.

14. The isolated nucleic acid fragment of Claim 1 wherein the isolated nucleic acid fragment is at least 80% similar to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in SEQ ID NO:10.

15. The isolated nucleic acid fragment of Claim 1 wherein the isolated nucleic acid fragment is at least 80% similar to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in SEQ ID NO:12.

16. The isolated nucleic acid fragment of Claim 1 wherein the isolated nucleic acid fragment is at least 80% similar to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in SEQ ID NO:14.

17. The isolated nucleic acid fragment of Claim 1 wherein the isolated nucleic acid fragment is at least 80% similar to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in SEQ ID NO:16.

18. The isolated nucleic acid fragment of Claim 1 wherein the nucleotide sequence of the fragment comprises all or a portion of the sequence set forth in a member selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQID NO:11, SEQ ID NO:13 and SEQ ID NO:15.

19. A chimeric gene comprising the nucleic acid fragment of Claim 1 operably linked to suitable regulatory sequences.

20. A transformed host cell comprising the chimeric gene of Claim 19.

21. An anthranilate synthase alpha subunit polypeptide comprising all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQID NO:12, SEQ ID NO:14 and SEQ ID NO:16.

22. An isolated nucleic acid fragment encoding all or a substantial portion of an anthranilate synthase beta subunit comprising a member selected from the group consisting of:

(a) an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22 and SEQ ID NO:24;

(b) an isolated nucleic acid fragment that is substantially similar to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22 and SEQ ID NO:24; and

(c) an isolated nucleic acid fragment that is complementary to (a) or (b).

23. The isolated nucleic acid fragment of Claim 22 wherein the isolated nucleic acid fragment hybridizes under stringent conditions to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in SEQ ID NO:18.

24. The isolated nucleic acid fragment of Claim 22 wherein the isolated nucleic acid fragment hybridizes under stringent conditions to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in SEQ ID NO:20.

25. The isolated nucleic acid fragment of Claim 22 wherein the isolated nucleic acid fragment hybridizes under stringent conditions to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in SEQ ID NO:22.

26. The isolated nucleic acid fragment of Claim 22 wherein the isolated nucleic acid fragment hybridizes under stringent conditions to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in SEQ ID NO:24.

27. The isolated nucleic acid fragment of Claim 22 wherein the isolated nucleic acid fragment is at least 80% similar to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in SEQ ID NO:18.

28. The isolated nucleic acid fragment of Claim 22 wherein the isolated nucleic acid fragment is at least 80% similar to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in SEQ ID NO:20.

29. The isolated nucleic acid fragment of Claim 22 wherein the isolated nucleic acid fragment is at least 80% similar to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in SEQ ID NO:22.

30. The isolated nucleic acid fragment of Claim 22 wherein the isolated nucleic acid fragment is at least 80% similar to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in SEQ ID NO:24.

31. The isolated nucleic acid fragment of Claim 22 wherein the nucleotide sequence of the fragment comprises all or a portion of the sequence set forth in a member

selected from the group consisting of SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21 and SEQ ID NO:23.

32. A chimeric gene comprising the nucleic acid fragment of Claim 22 operably linked to suitable regulatory sequences.

33. A transformed host cell comprising the chimeric gene of Claim 32.

34. An anthranilate synthase beta subunit polypeptide comprising all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22 and SEQ ID NO:24.

35. An isolated nucleic acid fragment encoding all or a substantial portion of a tryptophan synthase alpha subunit comprising a member selected from the group consisting of:

(a) an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36 and SEQ ID NO:38;

(b) an isolated nucleic acid fragment that is substantially similar to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36 and SEQ ID NO:38; and

(c) an isolated nucleic acid fragment that is complementary to (a) or (b).

36. The isolated nucleic acid fragment of Claim 35 wherein the isolated nucleic acid fragment hybridizes under stringent conditions to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in SEQ ID NO:26.

37. The isolated nucleic acid fragment of Claim 35 wherein the isolated nucleic acid fragment hybridizes under stringent conditions to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in SEQ ID NO:28.

38. The isolated nucleic acid fragment of Claim 35 wherein the isolated nucleic acid fragment hybridizes under stringent conditions to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in SEQ ID NO:30.

39. The isolated nucleic acid fragment of Claim 35 wherein the isolated nucleic acid fragment hybridizes under stringent conditions to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in SEQ ID NO:32.

40. The isolated nucleic acid fragment of Claim 35 wherein the isolated nucleic acid fragment hybridizes under stringent conditions to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in SEQ ID NO:34.

41. The isolated nucleic acid fragment of Claim 35 wherein the isolated nucleic acid fragment hybridizes under stringent conditions to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in SEQ ID NO:36.

42. The isolated nucleic acid fragment of Claim 35 wherein the isolated nucleic acid fragment hybridizes under stringent conditions to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in SEQ ID NO:38.

43. The isolated nucleic acid fragment of Claim 35 wherein the isolated nucleic acid fragment is at least 80% similar to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in SEQ ID NO:26.

44. The isolated nucleic acid fragment of Claim 35 wherein the isolated nucleic acid fragment is at least 80% similar to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in SEQ ID NO:28.

45. The isolated nucleic acid fragment of Claim 35 wherein the isolated nucleic acid fragment is at least 80% similar to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in SEQ ID NO:30.

46. The isolated nucleic acid fragment of Claim 35 wherein the isolated nucleic acid fragment is at least 80% similar to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in SEQ ID NO:32.

47. The isolated nucleic acid fragment of Claim 35 wherein the isolated nucleic acid fragment is at least 80% similar to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in SEQ ID NO:34.

48. The isolated nucleic acid fragment of Claim 35 wherein the isolated nucleic acid fragment is at least 80% similar to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in SEQ ID NO:36.

49. The isolated nucleic acid fragment of Claim 35 wherein the isolated nucleic acid fragment is at least 80% similar to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in SEQ ID NO:38.

50. The isolated nucleic acid fragment of Claim 35 wherein the nucleotide sequence of the fragment comprises all or a portion of the sequence set forth in a member selected from the group consisting of SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33, SEQ ID NO:35 and SEQ ID NO:37.

51. A chimeric gene comprising the nucleic acid fragment of Claim 35 operably linked to suitable regulatory sequences.

52. A transformed host cell comprising the chimeric gene of Claim 51.

53. A tryptophan synthase alpha subunit polypeptide comprising all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36 and SEQ ID NO:38.

54. A method of altering the level of expression of a tryptophan biosynthetic enzyme in a host cell comprising:

(a) transforming a host cell with the chimeric gene of any of Claims 19, 32 and 51; and

(b) growing the transformed host cell produced in step (a) under conditions that are suitable for expression of the chimeric gene

wherein expression of the chimeric gene results in production of altered levels of a tryptophan biosynthetic enzyme in the transformed host cell.

55. A method of obtaining a nucleic acid fragment encoding all or a substantial portion of the amino acid sequence encoding a tryptophan biosynthetic enzyme comprising:

(a) probing a cDNA or genomic library with the nucleic acid fragment of any of Claims 1, 22 and 35;

(b) identifying a DNA clone that hybridizes with the nucleic acid fragment of any of Claims 1, 22 and 35;

(c) isolating the DNA clone identified in step (b); and

(d) sequencing the cDNA or genomic fragment that comprises the clone isolated in step (c)

wherein the sequenced nucleic acid fragment encodes all or a substantial portion of the amino acid sequence encoding a tryptophan biosynthetic enzyme.

56. A method of obtaining a nucleic acid fragment encoding a substantial portion of an amino acid sequence encoding a tryptophan biosynthetic enzyme comprising:

(a) synthesizing an oligonucleotide primer corresponding to a portion of the sequence set forth in any of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35 and 37; and

(b) amplifying a cDNA insert present in a cloning vector using the oligonucleotide primer of step (a) and a primer representing sequences of the cloning vector

wherein the amplified nucleic acid fragment encodes a substantial portion of an amino acid sequence encoding a tryptophan biosynthetic enzyme.

57. The product of the method of Claim 55.

58. The product of the method of Claim 56.

59. A method for evaluating at least one compound for its ability to inhibit the activity of a tryptophan biosynthetic enzyme, the method comprising the steps of:

(a) transforming a host cell with a chimeric gene comprising a nucleic acid fragment encoding a tryptophan biosynthetic enzyme, operably linked to suitable regulatory sequences;

(b) growing the transformed host cell under conditions that are suitable for expression of the chimeric gene wherein expression of the chimeric gene results in production of the tryptophan biosynthetic enzyme encoded by the operably linked nucleic acid fragment in the transformed host cell;

(c) optionally purifying the tryptophan biosynthetic enzyme expressed by the transformed host cell;

(d) treating the tryptophan biosynthetic enzyme with a compound to be tested; and

(e) comparing the activity of the tryptophan biosynthetic enzyme that has been treated with a test compound to the activity of an untreated tryptophan biosynthetic enzyme,

thereby selecting compounds with potential for inhibitory activity.

FIG. 1

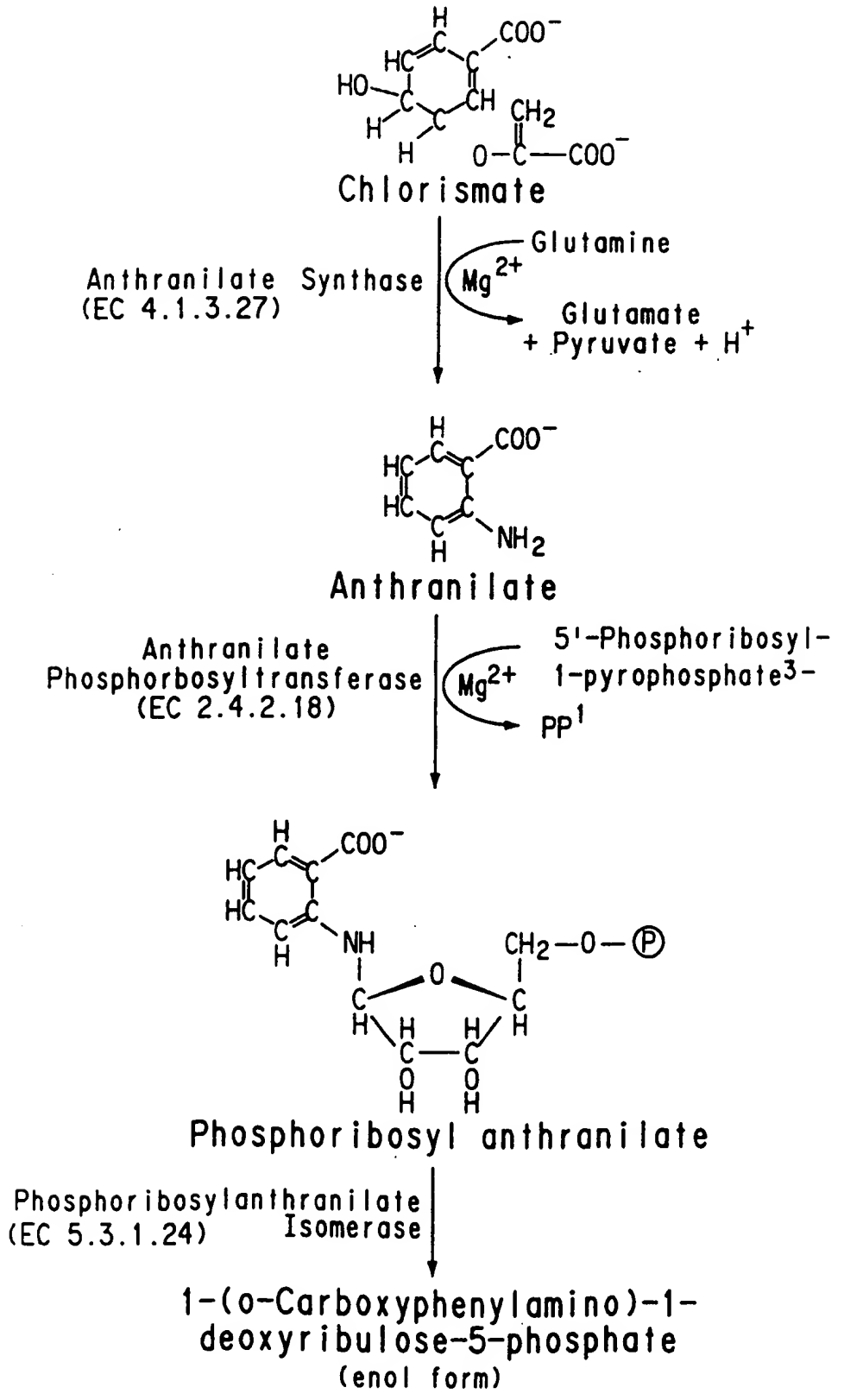
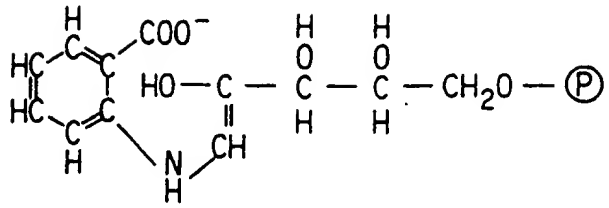


FIG. 1 (CONTINUED)

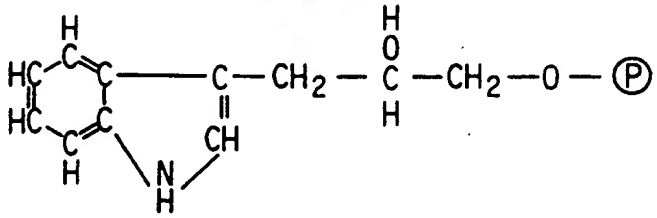


1-(o-Carboxyphenylamino)-1-deoxyribulose-5-phosphate

Indole-3-Glycol Phosphate Synthase (EC 4.1.1.20)

H^+

$CO_2 + H_2O$

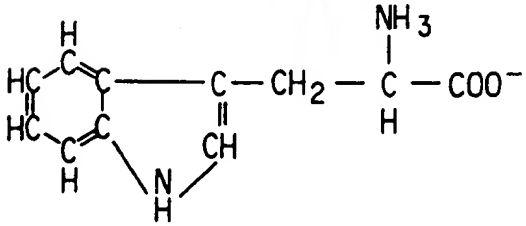


Indoleglycerol phosphate

Tryptophan synthase (EC 4.2.1.20)

Serine

Glyceraldehyde-3-phosphate



Tryptophan

Figure 2

```
                  *  *          *   *       *  *     * * *         * *       **
SEQ ID NO:39      MITLNVETP-PLTRSQLPSTFRVSSAASVNFNDRVATSRWRPNSLSLTTSSYRLRTLKCA
SEQ ID NO:02      MESLAATSVFAPSRXAVPAARALVRAGTV-----VPTRR--------TSSRSGTSGVKCS
                  1                                                        60

                  *     **    * *                        *       *           *   ***
SEQ ID NO:39      ASASTSASTSASPSPSPSLVDQSANFHEASKK-GNLIPLYRCIFSDH----LTPVLAYRC
SEQ ID NO:02      AAVTPQASPVISRSAAAAKAAEEDKRRXSRRRRGGTGRGTWCPCGSASCRTIYPRARYRC
                  61                                                       120

                  ** **   ********** * *     **** *** *  *****   *** ***  * **
SEQ ID NO:39      LVKEDDRDAPSFLFESVEPGSQASS-IGRYSVVGAQPAIEIVAKENMVTILDHEGGQRTE
SEQ ID NO:02      LVPEDNVXAPSFLFESVEQGPQGTTNVGRYSMVGAHPVMEIVAKDHKVTIMDHEKSQVTE
                  121                                                      180

                  * * ***  **  **** ** ****** * ***** ***************** *** ***
SEQ ID NO:39      QFVEDPMDVPRRIMEGWKPQLIDELPEAFCGGWVGYFSYDTVRYVEKKKLPFFSAPTDDR
SEQ ID NO:02      QVVDDPMQIPRTMMEGWHPQQIDELPESFSGGWVGFFSYDTVRYVEKKKLPFSSAPQDDR
                  181                                                      240

                  ************* *** ****  *****  *   ** *** **  **  * * **
SEQ ID NO:39      NLPDVHLGLYDDVIVFDHVEKKAFVIHWVRLDQYSSVAEAYNDGMNRLENLVSRVHDIVP
SEQ ID NO:02      NLPDVHLGLYDDVLVFDNVEKKVYVIHWVNVDRHASVEEAYQDGRSRLNMLLSKVHNSNV
                  241                                                      300
```

Figure 2 (Cont.)

```
                 * *  *  ***** **  *  * ***  ** *** ***** **  ************** *
SEQ ID NO:39     PKLRSGSIKLHTRHFGPKLERSSMTSEAYKEAVLEAKEHILAGDIFQIVLSQRFERRTFA
SEQ ID NO:02     PTLSPGFVKLHTRKFGTPLNKSTMTSDEYKNAVLQAKEHIMAGGSFQIVLSQRFERRTYA
                 301                                                       360

                  *** ** ********** * ***** ************ * ** ******* ***** *
SEQ ID NO:39     DPFEIYRSLRIVNPSPYMTYLQARGCILVASSPEILTRVKKRKITNRPLAGTIRRGKTRK
SEQ ID NO:02     NPFEVYRALRIVNPSPYMAYVQARGCVLVASSPEILTRVSKGKIINRPLAGTVRRGKTEK
                 361                                                       420

                 **   *  ** ********************** ************* ************
SEQ ID NO:39     EDLVFEKELLNDEKQCAEHIMLVDLGRNDVGKVSEPGSVKVEKLMNIEHYSHVMHISSTV
SEQ ID NO:02     EDQMQEQQLLSDEKQCAEHIMLVDLGRNDVGKVSKPGSVKVEKLMNIERYSHVMHISSTV
                 421                                                       480

                  * * *** ************************ ************ ** ***** **
SEQ ID NO:39     TGELLDHLTSWDALRAALPVGTVSGAPKVKAMEIIDKLEVTRRGPYGGGFGGISFTGDLD
SEQ ID NO:02     SGQLDDHLQSWDALRAALPVGTVSGAPKVKAMELIDKLEVTRRGPYSGGLGGISFDGDMQ
                 481                                                       540

                 *** *** ** **    ******* * **** ************* * ************
SEQ ID NO:39     IALALRTMVFQTATRYDTMYSYKDVDKRREWIAHLQAGAGIVADSDPADEQRECENKAAA
SEQ ID NO:02     IALSLRTIVFSTAPSHNTMYSYKDADRRREWVAHLQAGAGIVADSSPDDEQRECENKAAA
                 541                                                       600
```

Figure 2 (Cont.)

```
**********  *  *
SEQ ID NO:39  LARAIDLAESSFIEK.
SEQ ID NO:02  LARAIDLAESAFVDKE.
              601          617
```

Figure 3

```
                                                                *
SEQ ID NO:40   MAASTLYKSCLLQPKSGSTT-RRLNP-SLVNPLTNPTRVSVLG---KSRRDVFA-KASI
SEQ ID NO:20   MACSHIAAAGVSSPAAAAAR-SPAHSPASAFARLRSTPRFASAGLSVKGNGAAFPLVAAA
SEQ ID NO:22   ATPPLFSSDGHRRAAPPQDPVPRLPGRRGGAEARPSSLRLGVTSGPARTLKQKLVAKSAV
               1                                                          60

                                          ********************** *  *******  * *
SEQ ID NO:40   EMAESNSIPSVVVNSSKQHGPIIVIDNYDSFTYNLCQYMGELGCHFEVYRNDELTVEELK
SEQ ID NO:20   GPAAAAPVADLDGRPATEKQPIIVIDNYDSFTYNLCQYMGELGLNFEVYRNDELTIEDVR
SEQ ID NO:22   SVVEGENAFDGV---KQDTRPIIVIDNYDSFTYNLCQYMGEVGANFEVYRNDDITVEEIK
               61                                                         120

                  *** ******* ***********  ***  * ****************** * .* **
SEQ ID NO:40   KKNPRGVLISPGPGTPQDSGISLQTVLELGPLVPLFGVCMGLQCIGEAFGGKIVRSPFGV
SEQ ID NO:20   RKNPRGILISPGPGEPQDSGISLQTVLELGPTIPIFGVCMGLQCIGEAFGGKIIRAPSGV
SEQ ID NO:22   KISPRGILISPGPGTPQDSGISLQTVQDLGPSTPLFGVCMGLQCIGEAFGGKVVRSPYGV
               121                                                        180

                *** * * * *     ** ** ***   ********   ** * ** **** *** ***
SEQ ID NO:40   MHGKSSMVHYDEKGEEGLFSGLSNPFIVGRYHSLVIEKDTFPSDELEVTAWTEDGLVMAA
SEQ ID NO:20   MHGKSSPVYYDEELGKALFNGLPNPFTAARYHSLVIEQETFPHDALEATAWTEDGLIMAA
SEQ ID NO:22   VHGKGSLVHYEEKLDGTLFSGLPNPFQAGRYHSLVIEKDSFPHDALEITAWTDDGLIMAA
               181                                                        240
```

Figure 3..(Cont.)

```
                  *  *  *   *
SEQ ID NO:40  RHRKYKHIQGVQFHPESIITTEGKTIVRNFIKIVEKKESEKLT
SEQ ID NO:20  RHKKYKHIQGVQFHPESIITPEGKKIILNFVRFIEELEKQRS
SEQ ID NO:22  RTRKSKQYTG.....................................
                241                                    283
```

Figure 4

```
                                                                  *
SEQ ID NO:41  ............MAIAFKSGVFFLQSPKSQIGFRHSSPPDSSLSFKRFTPMASLST----
SEQ ID NO:26  ............MAFALKAA------AAGSASFSAAGP--------RRRAAAT-GRVSF-
SEQ ID NO:28  ....ARAQADRKLAIKAAPLRLRVRHSVHRPPPPRPPPRRLLLTPQRQRSEVALCRVEWW
SEQ ID NO:32  RTCSSRGHPLRRLHFVAVAV------AVAAAAASLSSAWAALL---RFRGGGG-AGLGLS
              1                                                          60

                                                      * ** *** **** **
SEQ ID NO:41  ......................SSPTLGLADTFTQLKKQGKVAFIPYITAGDPDLSTTA
SEQ ID NO:26  ---RS-AAPVVAVRAAAAAAAAVAEDKRSISGTFAELRQQGKTALIPFITAGDPDLATTA
SEQ ID NO:28  THPRPWPTAALRPASSPSPRPSPTSGSKARVGLLAPWWPIPNSAFIPFITAGDPDLVTTS
SEQ ID NO:32  ARDRE.PGRLPLMEMEDSGRGVVGAGKRGVAETFSRLREQGKTAFIPFITASDPDLATTS
              61                                                         120

                ** * **** ** ********** *** * * * * * **
SEQ ID NO:41  EALKVLDACGSDIIELGVPYSDPLADGPVIQAAATRSLERGTNLDSILEMLDKVVPQISC
SEQ ID NO:26  KALRILDACGSDVIELGVPYSDPLADGPVIQASATRALAKGTTFEDVISMVKGVIPDLSC
SEQ ID NO:28  KALKILNSCGSDVIEVGVPYSDPLADGPVIQASATRALKKGTTLDSVIEMLKGVTPELSC
SEQ ID NO:32  KALKILDSCGSDVIELGVPYSDPLADGPVIQAAATRALKKGATFDSVIAMLKGVIPELSC
              121                                                        180
```

Figure 4 (Cont.)

```
                  *   ***********   **      *  ****** *****  ** **     *******
SEQ ID NO:41      PISLFTYYNPILKRGLGKFMSSIRAVGVQGLVVPDVPLEETEMLRKEALNNDIELVLLTT
SEQ ID NO:26      PVALFTYYNPILKRGVPNFMSIVKEAGVHGLVVPDVPLEETDVLRSEAAKNNLELVLLTT
SEQ ID NO:28      PIVLFTYYNPILKRGVGNFMSTIKQAGIHGLVVPDLPLEETALLRSEAIMHNIELVLLTT
SEQ ID NO:32      PIVIFTYYNPILKRGVSNFMAIIKQAGVHGLVVPDLPLEETALLRNEAVMHGIELVLLTT
                  181                                                       240

                  ****  **  *  ** ** ****  ****    *   *  **   *  * ******** *
SEQ ID NO:41      PTTPTERMKLIVDASEGFIYLVSSIGVTGARSSVSGKVQSLLKDIKEATDKPVAVGFGIS
SEQ ID NO:26      PTTPNERMEKIAQASQGFIYLVSTVGVTGTPANVSGKVQSLLQDIKNVTEKPVAVGFGVS
SEQ ID NO:28      PTTPTDRMKGIAQASEGFLYLVSAVGVTGAPSNVNLRVEHLLRESKKVTDKPVAVGFGVS
SEQ ID NO:32      PTTPTERMKEIAKASEGFIYLVSSVGVTGARSNVNLRVEYLLQEIKKVTDKPVAVGFGIS
                  241                                                       300

                   **** ** ******** ***       *  * ***   *      * *
SEQ ID NO:41      KPEHVKQIAGWGADGVIVGSAMVKLLGDAKSPTEGLKELEKLTKSLKSALL.
SEQ ID NO:26      TPEHVRQIAGWGADGVIIGSAVMKTLEEAASPEEGLKKLEEFAKNLKAALP.
SEQ ID NO:28      TPEHVKQIVGWGADGVIVGSAIVKQLCEAATPEEGLERLEEYARSMKAAMP.
SEQ ID NO:32      TPEHVKQIAGWGADGVIIGSAIVRQLGEAASPEEGLKRVEEYAKNMKAAMP.
                  301                                                 352
```

SEQUENCE LISTING

<110> E. I. DU PONT DE NEMOURS AND COMPANY

<120> TRYPTOPHAN BIOSYNTHETIC ENZYMES

<130> BB-1150

<140>
<141>

<150> 60/079,386
<151> 1998-03-26

<160> 41

<170> Microsoft Word Version 7.0A

<210> 1
<211> 2266
<212> DNA
<213> Zea mays

<220>
<221> unsure
<222> (237)

<220>
<221> unsure
<222> (488)

<220>
<221> unsure
<222> (516)

<220>
<221> unsure
<222> (536)

<400> 1
ggttcttcct ccccttcac cgtaccgaga cccaaaccaa gccgcctcgc cagcacccag 60
cagcccgcat cactcgtctc cattaaaatc ggttttctcc ttggaaatct gctggcgagt 120
cccaggaccc gaatgcccca ctcttattac ctgctaattt tgaatttcct aatcgggttg 180
cagctgctta tataagctca ctaaccccgc attcgctttt ccgactccca cctccagcca 240
aatcggccta tagatcaaac gctgcactgt agggagcgtg aagccagcgg cgaatggaat 300
ccctagccgc cacctccgtg ttcgcgccct cccgcgncgc cgtcccggcg gcgcgggccc 360
tggttagggc ggggacggtg gtaccaacca ggcggacgag cagccggagc ggaaccagcg 420
gggtgaaatg ctctgctgcc gtgacgccgc aggcgagccc agtgattagc aggagcgctg 480
cggcggcnaa ggcggcggag gaggacaaga ggcggncttc gaggcggcgg cgcggnggaa 540
cgggaagggg aacctggtgc ccatgtggga gtgcatcgtg tcggaccatc tacccccgtg 600
ctcgctaccg ctgcctcgtc cccgaggaca acgtcnacgc ccccagcttc ctcttcgagt 660
ccgtcgagca ggggccccag ggcaccacca acgtcggccg ctacagcatg gtgggagccc 720
acccagtgat ggagattgtg gccaaagacc acaaggttac gatcatggac cacgagaaga 780
gccaagtgac agagcaggta gtggacgacc cgatgcagat cccgaggacc atgatggagg 840
gatggcaccc acagcagatc gacgagctcc ctgaatcctt ctccggtgga tgggttgggt 900
tcttttccta tgatacggtt aggtatgttg agaagaagaa gctaccgttc tccagtgctc 960
ctcaggacga taggaacctt cctgatgtgc acttgggact ctatgatgat gttctagtct 1020
tcgataatgt tgagaagaaa gtatatgtta tccattgggt caatgtggac cggcatgcat 1080
ctgttgagga agcataccaa gatggcaggt cccgactaaa catgttgcta tctaaagtgc 1140
acaattccaa tgtccccaca ctctctcctg gatttgtgaa gctgcacaca cgcaagtttg 1200
gtacaccttt gaacaagtcg accatgacaa gtgatgagta taagaatgct gttctgcagg 1260
ctaaggaaca tattatggct gggggaagct tccagattgt tttaagccag aggttcgaga 1320
gacgaacata tgccaaccca tttgaggttt atcgagcatt acggattgtg aatcctagcc 1380
catacatggc gtatgtacag gcaagaggct gtgtattggt tgcgtctagt cctgaaattc 1440
ttacacgagt cagtaagggg aagattatta atcgaccact tgctggaact gttcgaaggg 1500
gcaagacaga gaaggaagat caaatgcaag agcagcaact gttaagtgat gaaaaacagt 1560

```
gtgccgagca cataatgctt gtggacttgg gaaggaatga tgttggcaag gtatccaaac 1620
caggatcagt gaaggtggag aagttgatga acattgagag atactcccat gttatgcaca 1680
tcagctcaac ggttagtgga cagttggatg atcatctcca gagttgggat gccttgagag 1740
ctgccttgcc cgttggaaca gtcagtggtg caccaaaggt gaaggccatg gagttgattg 1800
ataagttgga agttacgagg cgaggaccat atagtggtgg tctaggagga atatcgtttg 1860
atggtgacat gcaaattgca ctttctctcc gcaccatcgt attctcaaca gcgccgagcc 1920
acaacacgat gtactcatac aaagacgcag ataggcgtcg ggagtgggtc gctcatcttc 1980
aggctggtgc aggcattgtt gccgacagta gcccagatga cgaacaacgt gaatgcgaga 2040
ataaggctgc tgcactagct cgggccatcg atcttgcaga gtcagctttt gtagacaaag 2100
aatagtgtgc tatggttatc gtttagttct tgttcatgtt tcttttaccc actttccgtt 2160
aaaaaaagat gtcattagtg ggtggagaaa agcaataaga ctgttctcta gagaaccgaa 2220
gaaatatgga aattgaggtt atggcttaaa aaaaaaaaaa aaaaaa                 2266
```

<210> 2
<211> 603
<212> PRT
<213> Zea mays

<220>
<221> UNSURE
<222> (15)

<220>
<221> UNSURE
<222> (75)

<220>
<221> UNSURE
<222> (115)

<400> 2

```
Met Glu Ser Leu Ala Ala Thr Ser Val Phe Ala Pro Ser Arg Xaa Ala
 1               5                  10                  15

Val Pro Ala Ala Arg Ala Leu Val Arg Ala Gly Thr Val Val Pro Thr
            20                  25                  30

Arg Arg Thr Ser Ser Arg Ser Gly Thr Ser Gly Val Lys Cys Ser Ala
        35                  40                  45

Ala Val Thr Pro Gln Ala Ser Pro Val Ile Ser Arg Ser Ala Ala Ala
    50                  55                  60

Ala Lys Ala Ala Glu Glu Asp Lys Arg Arg Xaa Ser Arg Arg Arg Arg
65                  70                  75                  80

Gly Gly Thr Gly Arg Gly Thr Trp Cys Pro Cys Gly Ser Ala Ser Cys
                85                  90                  95

Arg Thr Ile Tyr Pro Arg Ala Arg Tyr Arg Cys Leu Val Pro Glu Asp
            100                 105                 110

Asn Val Xaa Ala Pro Ser Phe Leu Phe Glu Ser Val Glu Gln Gly Pro
        115                 120                 125

Gln Gly Thr Thr Asn Val Gly Arg Tyr Ser Met Val Gly Ala His Pro
    130                 135                 140

Val Met Glu Ile Val Ala Lys Asp His Lys Val Thr Ile Met Asp His
145                 150                 155                 160

Glu Lys Ser Gln Val Thr Glu Gln Val Val Asp Asp Pro Met Gln Ile
                165                 170                 175

Pro Arg Thr Met Met Glu Gly Trp His Pro Gln Gln Ile Asp Glu Leu
            180                 185                 190
```

Pro Glu Ser Phe Ser Gly Gly Trp Val Gly Phe Phe Ser Tyr Asp Thr
195 200 205

Val Arg Tyr Val Glu Lys Lys Lys Leu Pro Phe Ser Ser Ala Pro Gln
210 215 220

Asp Asp Arg Asn Leu Pro Asp Val His Leu Gly Leu Tyr Asp Asp Val
225 230 235 240

Leu Val Phe Asp Asn Val Glu Lys Lys Val Tyr Val Ile His Trp Val
245 250 255

Asn Val Asp Arg His Ala Ser Val Glu Glu Ala Tyr Gln Asp Gly Arg
260 265 270

Ser Arg Leu Asn Met Leu Leu Ser Lys Val His Asn Ser Asn Val Pro
275 280 285

Thr Leu Ser Pro Gly Phe Val Lys Leu His Thr Arg Lys Phe Gly Thr
290 295 300

Pro Leu Asn Lys Ser Thr Met Thr Ser Asp Glu Tyr Lys Asn Ala Val
305 310 315 320

Leu Gln Ala Lys Glu His Ile Met Ala Gly Gly Ser Phe Gln Ile Val
325 330 335

Leu Ser Gln Arg Phe Glu Arg Arg Thr Tyr Ala Asn Pro Phe Glu Val
340 345 350

Tyr Arg Ala Leu Arg Ile Val Asn Pro Ser Pro Tyr Met Ala Tyr Val
355 360 365

Gln Ala Arg Gly Cys Val Leu Val Ala Ser Ser Pro Glu Ile Leu Thr
370 375 380

Arg Val Ser Lys Gly Lys Ile Ile Asn Arg Pro Leu Ala Gly Thr Val
385 390 395 400

Arg Arg Gly Lys Thr Glu Lys Glu Asp Gln Met Gln Glu Gln Gln Leu
405 410 415

Leu Ser Asp Glu Lys Gln Cys Ala Glu His Ile Met Leu Val Asp Leu
420 425 430

Gly Arg Asn Asp Val Gly Lys Val Ser Lys Pro Gly Ser Val Lys Val
435 440 445

Glu Lys Leu Met Asn Ile Glu Arg Tyr Ser His Val Met His Ile Ser
450 455 460

Ser Thr Val Ser Gly Gln Leu Asp Asp His Leu Gln Ser Trp Asp Ala
465 470 475 480

Leu Arg Ala Ala Leu Pro Val Gly Thr Val Ser Gly Ala Pro Lys Val
485 490 495

Lys Ala Met Glu Leu Ile Asp Lys Leu Glu Val Thr Arg Arg Gly Pro
500 505 510

Tyr Ser Gly Gly Leu Gly Gly Ile Ser Phe Asp Gly Asp Met Gln Ile
515 520 525

Ala Leu Ser Leu Arg Thr Ile Val Phe Ser Thr Ala Pro Ser His Asn
530 535 540

Thr Met Tyr Ser Tyr Lys Asp Ala Asp Arg Arg Arg Glu Trp Val Ala
545 550 555 560

His Leu Gln Ala Gly Ala Gly Ile Val Ala Asp Ser Ser Pro Asp Asp
565 570 575

Glu Gln Arg Glu Cys Glu Asn Lys Ala Ala Ala Leu Ala Arg Ala Ile
580 585 590

Asp Leu Ala Glu Ser Ala Phe Val Asp Lys Glu
595 600

<210> 3
<211> 1563
<212> DNA
<213> Zea mays

<220>
<221> unsure
<222> (116)

<400> 3
ggccaaggct aaccatgtga cgggtcatgg accatgagat gaagtcgagg agggagcact 60
tcgtgcctga tcccatgagg atccccagga caatcatgga gcagtggaac ccgcanattg 120
ctgacagcct ccctgatgca ttttgtggag gatgggttgg attcttctca tatgatacag 180
tgcgttatgt tgaaacaaag aagcttcctt tcagtaaggc accacatgac gataggaacc 240
ttcctgacat tcatttaggc ctctatagtg acgtcattgt gtttgatcat gttgaaaaga 300
aaacacatgt tattcattgg gtgaggacac actgctatcg ttctgttgat gaagcatatg 360
aagatggaag aaatcggctt gaagctttgt tatcaagatt acattgcctc aatgtcccaa 420
cactttcttc tggttctata aaactcaatg ttgaaaactt tggcccagta atgcaaaaat 480
caacgatgtc aagcgaagaa tataaaaata tcgttgtcca agctaaagaa cacatcttgg 540
ccggtgacat tttccaagtt gttttaagcc agcgttttga gagacggaca ttcgccgacc 600
cttttgaaat ctatcgtgca ttgcgcatcg taaatcctag tccatatatg gcctatctac 660
aggcacgagg ttgtattctc gtggcatcga gtcctgaaat tcttacccgg gtacaaaaga 720
ggacaataat caatcgtccg cttgctggaa ccataagaag aggcaaaaca aaagcagaag 780
acaaaacttt agaacaattg cttttgagtg atgaaaagca gtgtgctgaa catattatgc 840
tagtagatct tggccgaaat gatgttggga aggtgtccaa accaggttca gtaaaggtag 900
agaaattgat gaatatcgaa cgatattctc atgtcatgca catcagctca acagtaactg 960
gagagctacg cgatgatctt acgtgttggg atgcgctacg agccgcattg ccagttggaa 1020
ccgttagtgg cgctccaaag gtgagagcaa tggagttgat tgatcagcta gaagtgagta 1080
tgcgtgggcc gtatagtggt ggttttggag ggatttcctt ccgcggcgac atggacattg 1140
cactcgctct tcgcactatt gtcttcccca ccgcatctag gtttgatacc atgtactcgt 1200
acacagacag taagtcccgg caggagtggg tggctcacct ccaggccgga gctggcatag 1260
ttgctgatag caaaccggat gacgagcacc aagagtgtat aaacaaggct gcaggtgttg 1320
ctcgtgccat tgaccttgct gaatcgacat ttcttgaaga gtagtctagt ctaatgaagg 1380
aaatgtatgt ttaagttctc tgtataatta tggattgtcc tagaaaacag gctttcttag 1440
gccgaataaa aactcaactg taataaagtt aataaatgga caactttagc taaaaaaaaa 1500
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1560
aaa 1563

<210> 4
<211> 446
<212> PRT
<213> Zea mays

<220>
<221> UNSURE
<222> (31)

<400> 4
Val Met Asp His Glu Met Lys Ser Arg Arg Glu His Phe Val Pro Asp
1 5 10 15

Pro Met Arg Ile Pro Arg Thr Ile Met Glu Gln Trp Asn Pro Xaa Ile
20 25 30

```
Ala Asp Ser Leu Pro Asp Ala Phe Cys Gly Gly Trp Val Gly Phe Phe
        35                  40                  45

Ser Tyr Asp Thr Val Arg Tyr Val Glu Thr Lys Lys Leu Pro Phe Ser
    50                  55                  60

Lys Ala Pro His Asp Asp Arg Asn Leu Pro Asp Ile His Leu Gly Leu
65                  70                  75                  80

Tyr Ser Asp Val Ile Val Phe Asp His Val Glu Lys Lys Thr His Val
                85                  90                  95

Ile His Trp Val Arg Thr His Cys Tyr Arg Ser Val Asp Glu Ala Tyr
            100                 105                 110

Glu Asp Gly Arg Asn Arg Leu Glu Ala Leu Leu Ser Arg Leu His Cys
        115                 120                 125

Leu Asn Val Pro Thr Leu Ser Ser Gly Ser Ile Lys Leu Asn Val Glu
    130                 135                 140

Asn Phe Gly Pro Val Met Gln Lys Ser Thr Met Ser Ser Glu Glu Tyr
145                 150                 155                 160

Lys Asn Ile Val Val Gln Ala Lys Glu His Ile Leu Ala Gly Asp Ile
                165                 170                 175

Phe Gln Val Val Leu Ser Gln Arg Phe Glu Arg Arg Thr Phe Ala Asp
            180                 185                 190

Pro Phe Glu Ile Tyr Arg Ala Leu Arg Ile Val Asn Pro Ser Pro Tyr
        195                 200                 205

Met Ala Tyr Leu Gln Ala Arg Gly Cys Ile Leu Val Ala Ser Ser Pro
    210                 215                 220

Glu Ile Leu Thr Arg Val Gln Lys Arg Thr Ile Ile Asn Arg Pro Leu
225                 230                 235                 240

Ala Gly Thr Ile Arg Arg Gly Lys Thr Lys Ala Glu Asp Lys Thr Leu
                245                 250                 255

Glu Gln Leu Leu Leu Ser Asp Glu Lys Gln Cys Ala Glu His Ile Met
            260                 265                 270

Leu Val Asp Leu Gly Arg Asn Asp Val Gly Lys Val Ser Lys Pro Gly
        275                 280                 285

Ser Val Lys Val Glu Lys Leu Met Asn Ile Glu Arg Tyr Ser His Val
    290                 295                 300

Met His Ile Ser Ser Thr Val Thr Gly Glu Leu Arg Asp Asp Leu Thr
305                 310                 315                 320

Cys Trp Asp Ala Leu Arg Ala Ala Leu Pro Val Gly Thr Val Ser Gly
                325                 330                 335

Ala Pro Lys Val Arg Ala Met Glu Leu Ile Asp Gln Leu Glu Val Ser
            340                 345                 350

Met Arg Gly Pro Tyr Ser Gly Gly Phe Gly Gly Ile Ser Phe Arg Gly
        355                 360                 365

Asp Met Asp Ile Ala Leu Ala Leu Arg Thr Ile Val Phe Pro Thr Ala
    370                 375                 380
```

Ser Arg Phe Asp Thr Met Tyr Ser Tyr Thr Asp Ser Lys Ser Arg Gln
385 390 395 400

Glu Trp Val Ala His Leu Gln Ala Gly Ala Gly Ile Val Ala Asp Ser
405 410 415

Lys Pro Asp Asp Glu His Gln Glu Cys Ile Asn Lys Ala Ala Gly Val
420 425 430

Ala Arg Ala Ile Asp Leu Ala Glu Ser Thr Phe Leu Glu Glu
435 440 445

<210> 5
<211> 339
<212> DNA
<213> Oryza sativa

<220>
<221> unsure
<222> (132)

<220>
<221> unsure
<222> (147)

<220>
<221> unsure
<222> (246)

<220>
<221> unsure
<222> (251)

<220>
<221> unsure
<222> (326)

<220>
<221> unsure
<222> (332)

<400> 5
agatcagttc cacggtgagt ggagagttgg atgatcatct ccaaagttgg gatgccctgc 60
gagccgcgtt gcctgttgga acagttagtg gagcaccaaa ggtgaaagcc atggagctga 120
tagacgagct anaggtcaca agacgangac atacagtggc gggccttgga ggggatatca 180
ttttgacggg agacaatgct tatcgctcct tggcactccc gcaaccaatt tgtgttctca 240
acaagncgcc naagccaaca acaacgaatg ttctccatcc aaacgacaac cgagagggcc 300
ccggggaagt gggtcccctc aacttnaagg cnggtggct 339

<210> 6
<211> 57
<212> PRT
<213> Oryza sativa

<220>
<221> UNSURE
<222> (44)

<220>
<221> UNSURE
<222> (49)

<400> 6
Ile Ser Ser Thr Val Ser Gly Glu Leu Asp Asp His Leu Gln Ser Trp
1 5 10 15

Asp Ala Leu Arg Ala Ala Leu Pro Val Gly Thr Val Ser Gly Ala Pro
20 25 30

Lys Val Lys Ala Met Glu Leu Ile Asp Glu Leu Xaa Val Thr Arg Arg
35 40 45

Xaa His Thr Val Ala Gly Leu Gly Gly
50 55

<210> 7
<211> 513
<212> DNA
<213> Glycine max

<220>
<221> unsure
<222> (26)

<220>
<221> unsure
<222> (30)

<220>
<221> unsure
<222> (79)

<220>
<221> unsure
<222> (231)

<220>
<221> unsure
<222> (323)

<220>
<221> unsure
<222> (343)

<220>
<221> unsure
<222> (426)

<220>
<221> unsure
<222> (467)

<220>
<221> unsure
<222> (482)

<220>
<221> unsure
<222> (485)

<220>
<221> unsure
<222> (491)

<220>
<221> unsure
<222> (506)

<400> 7
atttctctta tgatacaatg cgctangtan aaaagaagaa acttccattt tctaatgccc 60
cagtagatga cagaaaccnt cctgatgttc atctgggcct ttatgacagt gtgattgtgt 120

ttgatcatgt tgaaaaagaa agcatatgtg attcattggg ttcgggtgga tcgatattct 180
tcaagctgag gaggccttcg aagatggaag gaaccggctg gaaactctag natctcgggt 240
gcatgatata attaccccaa ggctgcctac aggttcgata aagttataca ctcgtctctt 300
tggtcctaaa ctggagatgt canacatgac aaatgaggag tanaagaggg cagtattgaa 360
ggctaaagag cacaaaacgg gctgggtgat atttttcaaa attgtactaa agtcaaacgt 420
ttttgnacaa gagaactttt tgcaggaccc aatttgaaat ctacaanagg cattgagggg 480
antgnttaat ncccaagtcc catatnattg act 513

<210> 8
<211> 123
<212> PRT
<213> Glycine max

<220>
<221> UNSURE
<222> (8)

<220>
<221> UNSURE
<222> (10)

<220>
<221> UNSURE
<222> (26)

<220>
<221> UNSURE
<222> (76)

<220>
<221> UNSURE
<222> (107)

<220>
<221> UNSURE
<222> (113)

<400> 8
Phe Ser Tyr Asp Thr Met Arg Xaa Val Xaa Lys Lys Lys Leu Pro Phe
1 5 10 15

Ser Asn Ala Pro Val Asp Asp Arg Asn Xaa Pro Asp Val His Leu Gly
20 25 30

Leu Tyr Asp Ser Val Ile Val Phe Asp His Val Lys Lys Lys Ala Tyr
35 40 45

Val Ile His Trp Val Arg Val Asp Arg Tyr Ser Ser Ala Glu Glu Ala
50 55 60

Phe Glu Asp Gly Arg Asn Arg Leu Glu Thr Leu Xaa Ser Arg Val His
65 70 75 80

Asp Ile Ile Thr Pro Arg Leu Pro Thr Gly Ser Ile Lys Leu Tyr Thr
85 90 95

Arg Leu Phe Gly Pro Lys Leu Glu Met Ser Xaa Met Thr Asn Glu Glu
100 105 110

Xaa Lys Arg Ala Val Leu Lys Ala Lys Glu His
115 120

<210> 9
<211> 507
<212> DNA
<213> Glycine max

<220>
<221> unsure
<222> (405)

<220>
<221> unsure
<222> (414)

<220>
<221> unsure
<222> (479)

<220>
<221> unsure
<222> (485)..(486)

<400> 9
agaagaatga ccttttcctt ccttcgtcct ccaatggcga ctgttccgca cccattatcc 60
ctcgcaagtg taggttttgc taaccgaacc tcctccatct ccagatccac tctcaaatgc 120
tgcgctcaat ctccttctcc ttcactagtt gacaacgccc agaagtttct cgaagcttcc 180
aagaagggga acgtcattcc tctcttccgc tgcatatttt ccgatcacct cactccggtg 240
cttgcgtacc ggtgcctggt taaggaggac gagagagatg ctccgagttt tctctttgaa 300
tcggtcgagc caggccaaat ttctagcatc ggacggtaca gtgtggttgg agcacaagcc 360
gtgtatggga aattgtggcg aaagagaacg tgggttacta ttaanggaca acgntggaag 420
ggcgcaagga gtgaggaaat tgtaaaggtt cccccgggtg aattcctccg taaggggtna 480
atggnnaagt tggacgcctc aaatcct 507

<210> 10
<211> 130
<212> PRT
<213> Glycine max

<220>
<221> UNSURE
<222> (120)

<400> 10
Met Thr Phe Ser Phe Leu Arg Pro Pro Met Ala Thr Val Pro His Pro
1 5 10 15

Leu Ser Leu Ala Ser Val Gly Phe Ala Asn Arg Thr Ser Ser Ile Ser
20 25 30

Arg Ser Thr Leu Lys Cys Cys Ala Gln Ser Pro Ser Pro Ser Leu Val
35 40 45

Asp Asn Ala Gln Lys Phe Leu Glu Ala Ser Lys Lys Gly Asn Val Ile
50 55 60

Pro Leu Phe Arg Cys Ile Phe Ser Asp His Leu Thr Pro Val Leu Ala
65 70 75 80

Tyr Arg Cys Leu Val Lys Glu Asp Glu Arg Asp Ala Pro Ser Phe Leu
85 90 95

Phe Glu Ser Val Glu Pro Gly Gln Ile Ser Ser Ile Gly Arg Tyr Ser
100 105 110

Val Val Gly Ala Gln Ala Val Xaa Trp Glu Ile Val Ala Lys Glu Asn
115 120 125

Val Gly
130

<210> 11
<211> 452
<212> DNA
<213> Triticum aestivum

<400> 11
gcacgagcca gctccctgag gccttcagcg gtggatgggt tgggttcttt tcctatgata 60
cagtccgcta tgtcgaaaag aagaagatac ctttctctgg tgctccccac gatgatagga 120
accttcctga tgttcacttg gggctttacg atgatgttct tgtcttcgac aatgttgaga 180
agaaagtata tgtcatccat tgggtaagtg tggaccggca tgcatccacc gaggaagcat 240
acaaagatgg caggtcccgg ttgaagcggt tgctttctaa agttcacaat gcaaatgtcc 300
ccaagctctc tccaggattt gtgaagctac atactcggca gtttggtact cccttgaaca 360
aatcgaccat gacaagtgat gagtacaaga gtgctgttat gcaggctaag gagcatattc 420
tgggcggtaa tattttccag attgttttaa gc 452

<210> 12
<211> 150
<212> PRT
<213> Triticum aestivum

<400> 12
Thr Ser Gln Leu Pro Glu Ala Phe Ser Gly Gly Trp Val Gly Phe Phe
1 5 10 15

Ser Tyr Asp Thr Val Arg Tyr Val Glu Lys Lys Lys Ile Pro Phe Ser
20 25 30

Gly Ala Pro His Asp Asp Arg Asn Leu Pro Asp Val His Leu Gly Leu
35 40 45

Tyr Asp Asp Val Leu Val Phe Asp Asn Val Glu Lys Lys Val Tyr Val
50 55 60

Ile His Trp Val Ser Val Asp Arg His Ala Ser Thr Glu Glu Ala Tyr
65 70 75 80

Lys Asp Gly Arg Ser Arg Leu Lys Arg Leu Leu Ser Lys Val His Asn
85 90 95

Ala Asn Val Pro Lys Leu Ser Pro Gly Phe Val Lys Leu His Thr Arg
100 105 110

Gln Phe Gly Thr Pro Leu Asn Lys Ser Thr Met Thr Ser Asp Glu Tyr
115 120 125

Lys Ser Ala Val Met Gln Ala Lys Glu His Ile Leu Gly Gly Asn Ile
130 135 140

Phe Gln Ile Val Leu Ser
145 150

<210> 13
<211> 504
<212> DNA
<213> Triticum aestivum

<400> 13
cccaaacagt ggtggcttag gagggatatc atttgatggt gacatgctta tcgctcttgc 60
tctccgcacc attgtgtttt caacagctcc aagccccaat aggatgtact catacaaaag 120
ctcagatagg ccccgagagt gggttgctca tcttcaggct ggtgcgggca ttgttgctga 180
tagtatccca gacgatgagc aaaaagaatt tgagaataag gcggctgccc tagctcgggc 240
aattgatctt gcagagtcgg ctttttttaga caaagaatag agtgtctatt aaattatttt 300
ttttagttgt tcatcatttt tcacccagtt cattttggaa agttgttcat cgtttttttca 360
ccgagttcat attggggaaa aaaagcaata ccgttttgtt gtcctttgaa atgaataaat 420
ttgagctata ataagatgta ttttgctcat cgggcaaaaa aaaaaaaaaa aatataaaaa 480
aaaaaaaaaa aaaaaaaaaa aata 504

<210> 14
<211> 92
<212> PRT
<213> Triticum aestivum

<400> 14
Pro Asn Ser Gly Gly Leu Gly Gly Ile Ser Phe Asp Gly Asp Met Leu
1 5 10 15

Ile Ala Leu Ala Leu Arg Thr Ile Val Phe Ser Thr Ala Pro Ser Pro
20 25 30

Asn Arg Met Tyr Ser Tyr Lys Ser Ser Asp Arg Pro Arg Glu Trp Val
35 40 45

Ala His Leu Gln Ala Gly Ala Gly Ile Val Ala Asp Ser Ile Pro Asp
50 55 60

Asp Glu Gln Lys Glu Phe Glu Asn Lys Ala Ala Ala Leu Ala Arg Ala
65 70 75 80

Ile Asp Leu Ala Glu Ser Ala Phe Leu Asp Lys Glu
85 90

<210> 15
<211> 535
<212> DNA
<213> Triticum aestivum

<220>
<221> unsure
<222> (483)

<220>
<221> unsure
<222> (511)

<220>
<221> unsure
<222> (520)..(521)

<400> 15
ctgctttggg attcaactcc caccaccacc acctctaaac tcaaccggat tgcagcttct 60
catcccatct cgtatatata cgcgccctcc caatctcttt ctccgagcgc gaaatctgac 120
agagcagtgt cagaccgttg ttgagcagga agcggaagga cagttcaatc cgcaatggaa 180
tccctagccg ccgccacgtt ctcgccctcg cgcctcgctg cccaccccgc cccgcccgcg 240
gctgcggcag cgggcagatc gagggcggtg gcggcaacga ggaggaggag gagcagcagc 300
ggcctgaggt gctcgtcggc gagcgcgacc ccggtgatca atgggagcgc cgccgcgaag 360
gcggaggagg aggacaggag gcgcttcttt gaggcggcgg cacgggggac cggcaagggc 420
aacctggtgc ccgtgtggga gtgcatcgtg tcggattacc tcacccccgt gctcgcctaa 480
cgntgcctcg ttcccgaagg acgacattgg ncaaccccan nttcctcttc caagg 535

<210> 16
<211> 108
<212> PRT
<213> Triticum aestivum

<220>
<221> UNSURE
<222> (102)

<400> 16
Met Glu Ser Leu Ala Ala Ala Thr Phe Ser Pro Ser Arg Leu Ala Ala
1 5 10 15

```
His Pro Ala Pro Pro Ala Ala Ala Ala Ala Gly Arg Ser Arg Ala Val
            20                  25                  30

Ala Ala Thr Arg Arg Arg Arg Ser Ser Ser Gly Leu Arg Cys Ser Ser
        35                  40                  45

Ala Ser Ala Thr Pro Val Ile Asn Gly Ser Ala Ala Ala Lys Ala Glu
    50                  55                  60

Glu Glu Asp Arg Arg Arg Phe Phe Glu Ala Ala Ala Arg Gly Thr Gly
65                  70                  75                  80

Lys Gly Asn Leu Val Pro Val Trp Glu Cys Ile Val Ser Asp Tyr Leu
                85                  90                  95

Thr Pro Val Leu Ala Xaa Arg Cys Leu Val Pro Glu
            100                 105
```

<210> 17
<211> 797
<212> DNA
<213> Zea mays

<400> 17

```
gcacgaggaa attaagaaga tttctcctag aggaatactt atttccccag gccctggcac  60
tcctcaagat tcaggaatat ccttgcaaac agttacagag cttgggccct caataccgtt 120
gtttggggtt tgcatgggtt tacagtgcat tggagaggca tttggaggga aggttgttcg 180
atccccttac ggcgttgtgc acggaaaggg ctcccttgtt cattatgatg agaaacttga 240
tggaactctg ttttatgata ttcctaaccc attccaggct ggaaggtatc acagccttgt 300
aatcgagaag gatagtttcc cgcatgatac cctggaaatt gttgcgtgga cagatgatgg 360
gttgatcatg gctgctcgcc acaggatata caaacatatt cagggcgtgc agttccatcc 420
ggagagcatc ataactactg aagggaggct catggtcaac aattttatca agattattga 480
aggctacgaa gcctcgaatt gctctccgtg atttatctag ttggaagcag cagcagccct 540
gcgtggggcg ggtgacaata agatgatgta aaagcttgta actacatagc tcgaacatgt 600
aaactcagat gtatatgttc ggatgtcaca aactcacgac actcagagtc caactactgt 660
acatggtcgc atgtggcgtg tcgatgtcaa atcaaatcat tgaactttta gctcggtgac 720
cgctatattt gtttttttac agtagttacc actatatttg ttcatttggc aaaaaaaaaa 780
aaaaaaaaaa aaaaaaa                                               797
```

<210> 18
<211> 169
<212> PRT
<213> Zea mays

<400> 18

```
His Glu Glu Ile Lys Lys Ile Ser Pro Arg Gly Ile Leu Ile Ser Pro
  1               5                  10                  15

Gly Pro Gly Thr Pro Gln Asp Ser Gly Ile Ser Leu Gln Thr Val Thr
            20                  25                  30

Glu Leu Gly Pro Ser Ile Pro Leu Phe Gly Val Cys Met Gly Leu Gln
        35                  40                  45

Cys Ile Gly Glu Ala Phe Gly Gly Lys Val Val Arg Ser Pro Tyr Gly
    50                  55                  60

Val Val His Gly Lys Gly Ser Leu Val His Tyr Asp Glu Lys Leu Asp
65                  70                  75                  80

Gly Thr Leu Phe Tyr Asp Ile Pro Asn Pro Phe Gln Ala Gly Arg Tyr
                85                  90                  95

His Ser Leu Val Ile Glu Lys Asp Ser Phe Pro His Asp Thr Leu Glu
            100                 105                 110
```

```
Ile Val Ala Trp Thr Asp Asp Gly Leu Ile Met Ala Ala Arg His Arg
        115                 120                 125

Ile Tyr Lys His Ile Gln Gly Val Gln Phe His Pro Glu Ser Ile Ile
    130                 135                 140

Thr Thr Glu Gly Arg Leu Met Val Asn Asn Phe Ile Lys Ile Ile Glu
145                 150                 155                 160

Gly Tyr Glu Ala Ser Asn Cys Ser Pro
                165
```

<210> 19
<211> 1133
<212> DNA
<213> Zea mays

<220>
<221> unsure
<222> (13)

<400> 19

```
ctaccggttc cgnccagatg gcctgctccc acatcgccgc cgcgggggtc tcctcccccg   60
ccgcagcggc ggctcgttcc ccggcgcatt ctcccgcttc cgccttcgcg cgcctccggt  120
cgacgcctcg tttcgcgagc gctggcttgt cggttaaggg aaacggagcg gcgttcccgt  180
tggtcgccgc cgcgggggccg gccgcggcgg caccggtggc cgacctggac ggccgcccgg  240
caacggagaa gcagcccatc atcgtcatcg acaactacga cagcttcaca tacaacctct  300
gccagtatat gggggagctt ggattgaact tcgaagtata ccgcaatgat gaactgacca  360
tagaagatgt gagaaggaag aacccaaggg gaatacttat ttctccagga cctggtgaac  420
cacaagattc gggaatatca ttgcagactg ttcttgaatt aggcccaacc atcccaattt  480
ttggagtttg catgggtctg caatgcattg gagaggcatt tggaggaaag attatccgtg  540
ctccttctgg agtgatgcat gggaaaagct ctccagttta ttacgatgag gaattaggaa  600
aggcattgtt caatggcttg ccaaaccctt ttactgccgc gaggtaccac agcttggtca  660
ttgagcaaga aaccttccca catgatgctt tggaggctac tgcatggact gaagatggac  720
ttatcatggc tgctcgccac aagaagtaca aacacatcca gggtgtccaa ttccacccgg  780
agagcatcat caccctgaa ggcaagaaaa tcatcctcaa ctttgtcaga ttcattgagg  840
aactggagaa gcagcgttcg taggggaggt agacgcgaca ggtggtttca tttcatagat  900
cagacggaag cagagacaag gcgcttaaag ctgcgccagg ttccggggct ggcagttgaa  960
gcatagctag gaaacagctt tccctctctt aattcgttgt gttcgtggtg acataatctg 1020
tgtgcggacc gaatttagaa taaagtacag gcgtccgtct gatcaaataa gcaccggtgt 1080
tggtgctcta tagatgaata tatccggtga ttttcggcaa aaaaaaaaaa aaa        1133
```

<210> 20
<211> 281
<212> PRT
<213> Zea mays

<400> 20

```
Met Ala Cys Ser His Ile Ala Ala Ala Gly Val Ser Ser Pro Ala Ala
  1               5                  10                  15

Ala Ala Ala Arg Ser Pro Ala His Ser Pro Ala Ser Ala Phe Ala Arg
            20                  25                  30

Leu Arg Ser Thr Pro Arg Phe Ala Ser Ala Gly Leu Ser Val Lys Gly
        35                  40                  45

Asn Gly Ala Ala Phe Pro Leu Val Ala Ala Ala Gly Pro Ala Ala Ala
    50                  55                  60

Ala Pro Val Ala Asp Leu Asp Gly Arg Pro Ala Thr Glu Lys Gln Pro
65                  70                  75                  80

Ile Ile Val Ile Asp Asn Tyr Asp Ser Phe Thr Tyr Asn Leu Cys Gln
                85                  90                  95
```

```
Tyr Met Gly Glu Leu Gly Leu Asn Phe Glu Val Tyr Arg Asn Asp Glu
            100                 105                 110

Leu Thr Ile Glu Asp Val Arg Arg Lys Asn Pro Arg Gly Ile Leu Ile
        115                 120                 125

Ser Pro Gly Pro Gly Glu Pro Gln Asp Ser Gly Ile Ser Leu Gln Thr
    130                 135                 140

Val Leu Glu Leu Gly Pro Thr Ile Pro Ile Phe Gly Val Cys Met Gly
145                 150                 155                 160

Leu Gln Cys Ile Gly Glu Ala Phe Gly Gly Lys Ile Ile Arg Ala Pro
                165                 170                 175

Ser Gly Val Met His Gly Lys Ser Ser Pro Val Tyr Tyr Asp Glu Glu
            180                 185                 190

Leu Gly Lys Ala Leu Phe Asn Gly Leu Pro Asn Pro Phe Thr Ala Ala
        195                 200                 205

Arg Tyr His Ser Leu Val Ile Glu Gln Glu Thr Phe Pro His Asp Ala
    210                 215                 220

Leu Glu Ala Thr Ala Trp Thr Glu Asp Gly Leu Ile Met Ala Ala Arg
225                 230                 235                 240

His Lys Lys Tyr Lys His Ile Gln Gly Val Gln Phe His Pro Glu Ser
                245                 250                 255

Ile Ile Thr Pro Glu Gly Lys Lys Ile Ile Leu Asn Phe Val Arg Phe
            260                 265                 270

Ile Glu Glu Leu Glu Lys Gln Arg Ser
        275                 280
```

<210> 21
<211> 856
<212> DNA
<213> Oryza sativa

<220>
<221> unsure
<222> (811)

<220>
<221> unsure
<222> (816)..(816)

<220>
<221> unsure
<222> (831)

<220>
<221> unsure
<222> (836)

<400> 21

```
ttgccgttca gtcacagcaa aagaagccaa aaacttgatc aaatttctga gcaactccgc  60
cactcttctc ctccgatggc caccgccgcg cggctcctcc ccaagatcca gtcccccgcc 120
tccccggccg tcgcggaggc gcggaggcgc gcccctccag tctccgatta ggagttacta 180
gtggacccgc aagaactctg aagcaaaagc ttgttgctaa gagtgctgtt tctgtggtgg 240
aaggtgaaaa cgcatttgat ggagtaaagc aagatactag accaatcata gttatagata 300
actacgatag cttcacgtat aatttatgcc agtacatggg tgaggtggga gctaactttg 360
aggtgtaccg caatgattga tatcaccgtg gaagaaatta agaagatttc tcctagagga 420
atactcatct cccctggccc tggcacacct caagattcag gaatatcatt gcaaacagtt 480
```

caagatcttg gaccttctac acctttgttt ggggtttgca tgggtttgca gtgtattggg 540
gaggcatttg gagggaaggt tgttcgttct ccttatggag ttgtgcatgg gaaaggatcc 600
cttgttcact atgaggagaa acttgatgga acactgtttt ctggtctccc aaacccattc 660
caagcgggaa gataccacag ccttgtaatt gagaaggata gcttcccaca tgatgccctg 720
gaaattactg cttggacaga cgatgggctg atcatggctg ctcgcacaag gaagtccaaa 780
caatatacag ggtggtgcag tccatccaag ngagcnatca taacaccaga ngggangctc 840
aatggtccaa gaattt 856

<210> 22
<211> 247
<212> PRT
<213> Oryza sativa

<400> 22
Ala Thr Pro Pro Leu Phe Ser Ser Asp Gly His Arg Arg Ala Ala Pro
1 5 10 15

Pro Gln Asp Pro Val Pro Arg Leu Pro Gly Arg Arg Gly Gly Ala Glu
20 25 30

Ala Arg Pro Ser Ser Leu Arg Leu Gly Val Thr Ser Gly Pro Ala Arg
35 40 45

Thr Leu Lys Gln Lys Leu Val Ala Lys Ser Ala Val Ser Val Val Glu
50 55 60

Gly Glu Asn Ala Phe Asp Gly Val Lys Gln Asp Thr Arg Pro Ile Ile
65 70 75 80

Val Ile Asp Asn Tyr Asp Ser Phe Thr Tyr Asn Leu Cys Gln Tyr Met
85 90 95

Gly Glu Val Gly Ala Asn Phe Glu Val Tyr Arg Asn Asp Asp Ile Thr
100 105 110

Val Glu Glu Ile Lys Lys Ile Ser Pro Arg Gly Ile Leu Ile Ser Pro
115 120 125

Gly Pro Gly Thr Pro Gln Asp Ser Gly Ile Ser Leu Gln Thr Val Gln
130 135 140

Asp Leu Gly Pro Ser Thr Pro Leu Phe Gly Val Cys Met Gly Leu Gln
145 150 155 160

Cys Ile Gly Glu Ala Phe Gly Gly Lys Val Val Arg Ser Pro Tyr Gly
165 170 175

Val Val His Gly Lys Gly Ser Leu Val His Tyr Glu Glu Lys Leu Asp
180 185 190

Gly Thr Leu Phe Ser Gly Leu Pro Asn Pro Phe Gln Ala Gly Arg Tyr
195 200 205

His Ser Leu Val Ile Glu Lys Asp Ser Phe Pro His Asp Ala Leu Glu
210 215 220

Ile Thr Ala Trp Thr Asp Asp Gly Leu Ile Met Ala Ala Arg Thr Arg
225 230 235 240

Lys Ser Lys Gln Tyr Thr Gly
245

<210> 23
<211> 468
<212> DNA
<213> Triticum aestivum

<220>
<221> unsure
<222> (7)

<400> 23
ggaaccnttc ccgcacgttg aactggaagt cacggcatgg actgaagatg gacttgtcat 60
ggctgctcgc ccacaagaag tacacgcata tccagggtgt tcagttccca cccagagagc 120
atcatcaccc ctgaaggcaa gaaaatcatc ctcaactttg cgagatatgt cgaggagttt 180
gagaagcaga cctccgaggg gaagtagagg attcaggaag cagagcagtt agaggaagta 240
gagactattg ttgagagatc aggcaggcag gaaggggcct aaaagccaag cagatttagg 300
ggaaatgaat aagggaaaaa gagcttttcc attgccattc gtttggttca ttgtaatata 360
tcattgtgga tctagatccc gtatgaactt cattcggtca attatgttcg tcgtggtgct 420
ctttctcaaa taataaatcg gtggaagcaa aaaaaaaaaa aaaactcg 468

<210> 24
<211> 67
<212> PRT
<213> Triticum aestivum

<400> 24
Glu Pro Phe Pro His Val Glu Leu Glu Val Thr Ala Trp Thr Glu Asp
1 5 10 15

Gly Leu Val Met Ala Ala Arg His Lys Lys Tyr Thr His Ile Gln Gly
20 25 30

Val Gln Phe His Pro Glu Ser Ile Ile Thr Pro Glu Gly Lys Lys Ile
35 40 45

Ile Leu Asn Phe Ala Arg Tyr Val Glu Glu Phe Glu Lys Gln Thr Ser
50 55 60

Glu Gly Lys
65

<210> 25
<211> 1181
<212> DNA
<213> Zea mays

<400> 25
gcacgagctg aatccaccac ccacccacac gacaaaaatc cgccatggcc ttcgcgctca 60
aggccgccgc cgcggggtcg gcgtcgttct ccgccgcggg gccgaggcgg cgcgccgcgg 120
cgactggtag ggtctcgttc cggagcgccg cgcccgtggt cgccgtcagg gccgcggcgg 180
cggcggccgc cgctgtggcg gaggacaagc gcagcatctc cggcaccttc gccgagctca 240
ggcagcaggg gaagactgcc ttgattccct tcatcactgc tggagaccct gacttggcca 300
ccacagcaaa agcactcagg atccttgatg cgtgtggttc agacgtgatc gaactgggtg 360
tgccttactc cgatccattg gctgacggcc ctgttattca ggcctccgcg acgcgcgctc 420
tggcaaaggg caccacattt gaggatgtca tctccatggt aaaggggtg atacctgatc 480
tgtcctgccc tgtagcgctt ttcacatatt acaacccgat cctgaagcgt ggtgtcccca 540
acttcatgag tattgttaaa gaagctgggg tacacggtct tgtggtacca gatgttcctt 600
tggaagagac agatgtcttg aggagtgagg ctgccaaaaa caacctagag ctggtgctat 660
tgacaacacc gactacacca aatgaaagaa tggagaaaat tgcacaggct tcacaaggat 720
ttatttatct tgtaagcact gttggagtta ctggtacacc tgcaaatgta agcggcaagg 780
tgcaatctct tcttcaggat atcaagaatg tgacagaaaa acccgtggct gttggctttg 840
gtgtgtcgac tccagagcat gtccggcaga ttgcaggatg gggcgcagat ggtgtgatta 900
tcggtagcgc agtgatgaag acgttggagg aagctgcttc tccagaagaa ggattgaaga 960
agctagaaga gttcgccaag aacctgaagg ccgcattgcc gtgaaccaac cattttcgtt 1020
agcaaagaaa ttaagatttt tttttccttt caatcttggc tttcagggtt gttggttgtt 1080
gtgtagtgaa aataattgtg tccctatgtt gaatgctcaa atgaataaat caagatccca 1140
taaaaaaaaa aaaaaaaaac tcgaggggg gcccgtaccc a 1181

<210> 26
<211> 319

<212> PRT
<213> Zea mays

<400> 26
Met Ala Phe Ala Leu Lys Ala Ala Ala Ala Gly Ser Ala Ser Phe Ser
1 5 10 15

Ala Ala Gly Pro Arg Arg Arg Ala Ala Ala Thr Gly Arg Val Ser Phe
20 25 30

Arg Ser Ala Ala Pro Val Val Ala Val Arg Ala Ala Ala Ala Ala Ala
35 40 45

Ala Ala Val Ala Glu Asp Lys Arg Ser Ile Ser Gly Thr Phe Ala Glu
50 55 60

Leu Arg Gln Gln Gly Lys Thr Ala Leu Ile Pro Phe Ile Thr Ala Gly
65 70 75 80

Asp Pro Asp Leu Ala Thr Thr Ala Lys Ala Leu Arg Ile Leu Asp Ala
85 90 95

Cys Gly Ser Asp Val Ile Glu Leu Gly Val Pro Tyr Ser Asp Pro Leu
100 105 110

Ala Asp Gly Pro Val Ile Gln Ala Ser Ala Thr Arg Ala Leu Ala Lys
115 120 125

Gly Thr Thr Phe Glu Asp Val Ile Ser Met Val Lys Gly Val Ile Pro
130 135 140

Asp Leu Ser Cys Pro Val Ala Leu Phe Thr Tyr Tyr Asn Pro Ile Leu
145 150 155 160

Lys Arg Gly Val Pro Asn Phe Met Ser Ile Val Lys Glu Ala Gly Val
165 170 175

His Gly Leu Val Val Pro Asp Val Pro Leu Glu Glu Thr Asp Val Leu
180 185 190

Arg Ser Glu Ala Ala Lys Asn Asn Leu Glu Leu Val Leu Leu Thr Thr
195 200 205

Pro Thr Thr Pro Asn Glu Arg Met Glu Lys Ile Ala Gln Ala Ser Gln
210 215 220

Gly Phe Ile Tyr Leu Val Ser Thr Val Gly Val Thr Gly Thr Pro Ala
225 230 235 240

Asn Val Ser Gly Lys Val Gln Ser Leu Leu Gln Asp Ile Lys Asn Val
245 250 255

Thr Glu Lys Pro Val Ala Val Gly Phe Gly Val Ser Thr Pro Glu His
260 265 270

Val Arg Gln Ile Ala Gly Trp Gly Ala Asp Gly Val Ile Ile Gly Ser
275 280 285

Ala Val Met Lys Thr Leu Glu Glu Ala Ala Ser Pro Glu Glu Gly Leu
290 295 300

Lys Lys Leu Glu Glu Phe Ala Lys Asn Leu Lys Ala Ala Leu Pro
305 310 315

<210> 27
<211> 1243

<212> DNA
<213> Zea mays

<400> 27
gcacgagcgc aagcggatcg caaactcgca atcaaggctg ctcccctccg tctccgtgtt 60
cgccactcgg ttcaccgccc gccgccgccg agacctcctc ctcgccggct actcctgacc 120
ccacaaaggc agcgctcgga ggtggccctc tgtcgggtgg agtggtggac acatccgcgc 180
ccatggccaa cggcggcgct gcggccggca agctcaccgt cgccgagacc ttctccaacc 240
tcagggagca aggcaagggt tggattgctg gcgccgtggt ggccgatacc caatagtgca 300
ttcatcccgt tcataactgc tggtgatcct gacctggtaa ccacatcgaa agcattgaag 360
attctcaatt cctgcggttc ggatgttata gaggttggcg taccatactc ggatccactg 420
gctgatgggc cagttattca ggcttctgca acacgggcgc ttaagaaagg cactacactt 480
gattctgtca tagagatgct gaagggggtt acacctgagc tttcttgccc cattgttctt 540
ttcacatact ataatccaat tctgaaacgt ggtgtgggaa acttcatgtc tactatcaaa 600
caagctggca tacatggact tgtagtgcct gatcttcctt tggaggagac agcgttgctg 660
aggagcgagg ccattatgca caacatagag ctggtgctgc ttacaacgcc aaccacgccg 720
acagatcgga tgaagggaat tgcacaagct tcagaaggat ttctctatct tgtgagcgct 780
gtaggggtta caggtgcacg ctcaaatgta aacttacgtg tagaacacct tcttagggag 840
agcaagaagg ttacagacaa accggtagct gttggtttcg gtgtatcgac tccagagcat 900
gtgaaacaga tcgtgggctg gggagcggac ggtgtgattg ttggcagtgc gattgtgaag 960
cagttgtgtg aggctgctac tcctgaagag ggtctggaaa ggctcgagga gtacgccagg 1020
agtatgaagg ctgctatgcc gtgagctgat ctcactcccc gtcacaggca aggtaattag 1080
aacatggttt tagcaatatt ctgggcatgg ggttggcatc cataactcat ctagtgctgt 1140
acgggtcgtt ctccggtggt gagaacgaag tctggctttg cctataatag aaaagggtta 1200
taaactgagt tagtgcgaaa aaaaagataa gccttaaaaa aaa 1243

<210> 28
<211> 347
<212> PRT
<213> Zea mays

<400> 28
Ala Arg Ala Gln Ala Asp Arg Lys Leu Ala Ile Lys Ala Ala Pro Leu
1 5 10 15

Arg Leu Arg Val Arg His Ser Val His Arg Pro Pro Pro Pro Arg Pro
20 25 30

Pro Pro Arg Arg Leu Leu Leu Thr Pro Gln Arg Gln Arg Ser Glu Val
35 40 45

Ala Leu Cys Arg Val Glu Trp Trp Thr His Pro Arg Pro Trp Pro Thr
50 55 60

Ala Ala Leu Arg Pro Ala Ser Ser Pro Ser Pro Arg Pro Ser Pro Thr
65 70 75 80

Ser Gly Ser Lys Ala Arg Val Gly Leu Leu Ala Pro Trp Trp Pro Ile
85 90 95

Pro Asn Ser Ala Phe Ile Pro Phe Ile Thr Ala Gly Asp Pro Asp Leu
100 105 110

Val Thr Thr Ser Lys Ala Leu Lys Ile Leu Asn Ser Cys Gly Ser Asp
115 120 125

Val Ile Glu Val Gly Val Pro Tyr Ser Asp Pro Leu Ala Asp Gly Pro
130 135 140

Val Ile Gln Ala Ser Ala Thr Arg Ala Leu Lys Lys Gly Thr Thr Leu
145 150 155 160

Asp Ser Val Ile Glu Met Leu Lys Gly Val Thr Pro Glu Leu Ser Cys
165 170 175

```
Pro Ile Val Leu Phe Thr Tyr Tyr Asn Pro Ile Leu Lys Arg Gly Val
            180                 185                 190

Gly Asn Phe Met Ser Thr Ile Lys Gln Ala Gly Ile His Gly Leu Val
        195                 200                 205

Val Pro Asp Leu Pro Leu Glu Glu Thr Ala Leu Leu Arg Ser Glu Ala
    210                 215                 220

Ile Met His Asn Ile Glu Leu Val Leu Leu Thr Thr Pro Thr Thr Pro
225                 230                 235                 240

Thr Asp Arg Met Lys Gly Ile Ala Gln Ala Ser Glu Gly Phe Leu Tyr
                245                 250                 255

Leu Val Ser Ala Val Gly Val Thr Gly Ala Pro Ser Asn Val Asn Leu
            260                 265                 270

Arg Val Glu His Leu Leu Arg Glu Ser Lys Lys Val Thr Asp Lys Pro
        275                 280                 285

Val Ala Val Gly Phe Gly Val Ser Thr Pro Glu His Val Lys Gln Ile
    290                 295                 300

Val Gly Trp Gly Ala Asp Gly Val Ile Val Gly Ser Ala Ile Val Lys
305                 310                 315                 320

Gln Leu Cys Glu Ala Ala Thr Pro Glu Glu Gly Leu Glu Arg Leu Glu
                325                 330                 335

Glu Tyr Ala Arg Ser Met Lys Ala Ala Met Pro
            340                 345
```

<210> 29
<211> 831
<212> DNA
<213> Oryza sativa

<400> 29

```
tcctgccctg tagcgctttt cacatattat aacccgattc tgaagcgtgg catcgcaaac  60
ttcatgactg ttgtaaaaga agccggtgtc catggtcttg tggtacctga tgttcctctg 120
gaagagacaa atattttgag gagtgaagct gctaagaaca accttgagct ggtgctactg 180
acaacaccaa ctacaccaac agaaagaatg gagaaaatta caaaagcttc tgaaggattt 240
atttatcttg taagcactgt tggagttaca ggtgcacgtg caaatgtcag tggcaaggtg 300
caatctcttc tccaggatat caagcaggtc acggacaaag cttgtggctg ttgggttcgg 360
tatatcgact ccagagcatg tgaagcagat tgcgggatgg ggcgcagatg gtgtgatcat 420
tgggagcgct atggtgaggc aattgggcga agctgcttca cccgaagaag gattgaagaa 480
gctagaagag ctagccaaga gcctgaaggc tgcattgccc tgaaacaatc agttccatga 540
ggaaagaaaa tacagtccat ttttttgtag tcaggtttga agcatagtat tgggagcatg 600
acttttgcta ttggggtttt cggggttact gttgtctgtt gaaaataatt aagtcattgt 660
gttagttgct tgaacgaata aatcaaggta agaatgatac taaatgttgc gaaaagggcc 720
taacactgtt ttcgctttgt cgttggtgca aatatcgaag agcggatata aacggagaaa 780
ataatgtcgt aattcaccta aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa a          831
```

<210> 30
<211> 179
<212> PRT
<213> Oryza sativa

<400> 30

```
Ser Cys Pro Val Ala Leu Phe Thr Tyr Tyr Asn Pro Ile Leu Lys Arg
  1               5                  10                  15

Gly Ile Ala Asn Phe Met Thr Val Val Lys Glu Ala Gly Val His Gly
             20                  25                  30
```

```
Leu Val Val Pro Asp Val Pro Leu Glu Glu Thr Asn Ile Leu Arg Ser
        35                  40                  45

Glu Ala Ala Lys Asn Asn Leu Glu Leu Val Leu Leu Thr Thr Pro Thr
    50                  55                  60

Thr Pro Thr Glu Arg Met Glu Lys Ile Thr Lys Ala Ser Glu Gly Phe
65                  70                  75                  80

Ile Tyr Leu Val Ser Thr Val Gly Val Thr Gly Ala Arg Ala Asn Val
                85                  90                  95

Ser Gly Lys Val Gln Ser Leu Leu Gln Asp Ile Lys Gln Val Thr Asp
            100                 105                 110

Lys Ala Cys Gly Cys Trp Val Arg Tyr Ile Asp Ser Arg Ala Cys Glu
        115                 120                 125

Ala Asp Cys Gly Met Gly Arg Arg Trp Cys Asp His Trp Glu Arg Tyr
    130                 135                 140

Gly Glu Ala Ile Gly Arg Ser Cys Phe Thr Arg Arg Arg Ile Glu Glu
145                 150                 155                 160

Ala Arg Arg Ala Ser Gln Glu Pro Glu Gly Cys Ile Ala Leu Lys Gln
                165                 170                 175

Ser Val Pro
```

<210> 31
<211> 1226
<212> DNA
<213> Oryza sativa

<400> 31

```
cgcgcacgtg ctcatcgcga ggccacccgc tccgccgcct ccatttcgtc gccgtcgccg   60
ttgccgtcgc cgccgcggcg gcgtctctct cctccgcgtg ggctgctctg cttcgtttcc  120
gcggcggagg cggggcgggg ctgggcctgt cggccaggga cagggagcca gggaggttgc  180
cgctcatgga gatggaagat agcggtcgcg gcgtcgtggg cgccgggaag cgcggcgtcg  240
cggagacctt ctccaggctc cgggagcagg gcaagactgc attcattcca ttcatcaccg  300
ccagtgaccc tgacttggca accacatcga aagcgttgaa gatccttgat tcctgtggtt  360
cagatgtgat tgagctgggt gtaccttact cggatccgtt ggctgatggg ccggttattc  420
aggctgcagc aacgcgtgct cttaagaaag gcgctacatt tgattccgtc atagctatgc  480
tcaaggggt tatacctgag ttgtcttgtc caatagttat cttcacatac tacaacccaa  540
ttttaaagcg cggagtgagt aacttcatgg ctatcattaa acaagccggc gtacatggtc  600
ttgtagtacc tgatcttcct ttggaagaga cagcgctgtt gaggaatgag gctgtcatgc  660
atggcataga gctggtgctg cttacaacac caaccacacc aacagagagg atgaaggaga  720
ttgcaaaagc ttcagaagga tttatttatc ttgtaagttc tgttggagtt acaggtgcac  780
gctcaaatgt aaatttgcgt gttgagtacc ttcttcagga gatcaagaag gttacagaca  840
aacctgttgc tgttggcttt ggcatatcga ctccagagca tgtaaagcag attgcaggct  900
ggggagcaga tggtgtgatc atcggcagcg ctattgttag acagttagga gaggctgctt  960
cacctgaaga agggttgaaa agagtagaag agtatgccaa gaacatgaag gctgctatgc 1020
cttgagatga gcggttccat gcctccaccc atgcaagcca acaagaaggc agtacatcgt 1080
ttttttttgt ctgattaaat tgcatagcat tttaatccag tattttcttg tatgattcat 1140
gaatcatgac tgtatctcac tctagtacta tccttctgtt aacatacaga tgaataataa 1200
agaaggtagc tcagaaaaaa aaaaaa                                      1226
```

<210> 32
<211> 340
<212> PRT
<213> Oryza sativa

<400> 32

```
Arg Thr Cys Ser Ser Arg Gly His Pro Leu Arg Arg Leu His Phe Val
  1               5                  10                  15
```

```
Ala Val Ala Val Ala Val Ala Ala Ala Ala Ala Ser Leu Ser Ser Ala
            20                  25                  30

Trp Ala Ala Leu Leu Arg Phe Arg Gly Gly Gly Gly Ala Gly Leu Gly
        35                  40                  45

Leu Ser Ala Arg Asp Arg Glu Pro Gly Arg Leu Pro Leu Met Glu Met
    50                  55                  60

Glu Asp Ser Gly Arg Gly Val Val Gly Ala Gly Lys Arg Gly Val Ala
65                  70                  75                  80

Glu Thr Phe Ser Arg Leu Arg Glu Gln Gly Lys Thr Ala Phe Ile Pro
                85                  90                  95

Phe Ile Thr Ala Ser Asp Pro Asp Leu Ala Thr Thr Ser Lys Ala Leu
            100                 105                 110

Lys Ile Leu Asp Ser Cys Gly Ser Asp Val Ile Glu Leu Gly Val Pro
        115                 120                 125

Tyr Ser Asp Pro Leu Ala Asp Gly Pro Val Ile Gln Ala Ala Ala Thr
    130                 135                 140

Arg Ala Leu Lys Lys Gly Ala Thr Phe Asp Ser Val Ile Ala Met Leu
145                 150                 155                 160

Lys Gly Val Ile Pro Glu Leu Ser Cys Pro Ile Val Ile Phe Thr Tyr
                165                 170                 175

Tyr Asn Pro Ile Leu Lys Arg Gly Val Ser Asn Phe Met Ala Ile Ile
            180                 185                 190

Lys Gln Ala Gly Val His Gly Leu Val Val Pro Asp Leu Pro Leu Glu
        195                 200                 205

Glu Thr Ala Leu Leu Arg Asn Glu Ala Val Met His Gly Ile Glu Leu
    210                 215                 220

Val Leu Leu Thr Thr Pro Thr Thr Pro Thr Glu Arg Met Lys Glu Ile
225                 230                 235                 240

Ala Lys Ala Ser Glu Gly Phe Ile Tyr Leu Val Ser Ser Val Gly Val
                245                 250                 255

Thr Gly Ala Arg Ser Asn Val Asn Leu Arg Val Glu Tyr Leu Leu Gln
            260                 265                 270

Glu Ile Lys Lys Val Thr Asp Lys Pro Val Ala Val Gly Phe Gly Ile
        275                 280                 285

Ser Thr Pro Glu His Val Lys Gln Ile Ala Gly Trp Gly Ala Asp Gly
    290                 295                 300

Val Ile Ile Gly Ser Ala Ile Val Arg Gln Leu Gly Glu Ala Ala Ser
305                 310                 315                 320

Pro Glu Glu Gly Leu Lys Arg Val Glu Glu Tyr Ala Lys Asn Met Lys
                325                 330                 335

Ala Ala Met Pro
            340
```

<210> 33
<211> 533

<212> DNA
<213> Glycine max

<220>
<221> unsure
<222> (500)

<400> 33
gagagaagaa acccaataag tgattctcat ttttatggct cttgccctga agtcaagttg 60
cttcttgcaa ttgaagaaac ctgaggctgg tttcaatgtt tgtttttcat ctaagaaagc 120
tataatctcc gtgaaaagat acactccggt ggccgctatc aggatcatgg acgcagtggg 180
actctctgca acattcacca gattgaaaaa agaaggaaaa gtggctttca tcccatatat 240
cactgctggt gatcctgacc tttcaaccac tgcagaagca ctgaaattgc ttgattcatg 300
cgggtctgac ataattgagc taggcattcc atactctgat cctttggcag atggtcctgt 360
tatccaggct gctgctacaa gatctttagc aaaggggacc aattttaatg tgattattga 420
tatgttgaag gaggttgttc cgcaactgtc atgtccaatt gcactgttta cataattaca 480
atccaaatac ctggaaagcn gtgggtactg gagaaaattt atgtccaacc aat 533

<210> 34
<211> 146
<212> PRT
<213> Glycine max

<400> 34
Met Ala Leu Ala Leu Lys Ser Ser Cys Phe Leu Gln Leu Lys Lys Pro
1 5 10 15

Glu Ala Gly Phe Asn Val Cys Phe Ser Ser Lys Lys Ala Ile Ile Ser
20 25 30

Val Lys Arg Tyr Thr Pro Val Ala Ala Ile Arg Ile Met Asp Ala Val
35 40 45

Gly Leu Ser Ala Thr Phe Thr Arg Leu Lys Lys Glu Gly Lys Val Ala
50 55 60

Phe Ile Pro Tyr Ile Thr Ala Gly Asp Pro Asp Leu Ser Thr Thr Ala
65 70 75 80

Glu Ala Leu Lys Leu Leu Asp Ser Cys Gly Ser Asp Ile Ile Glu Leu
85 90 95

Gly Ile Pro Tyr Ser Asp Pro Leu Ala Asp Gly Pro Val Ile Gln Ala
100 105 110

Ala Ala Thr Arg Ser Leu Ala Lys Gly Thr Asn Phe Asn Val Ile Ile
115 120 125

Asp Met Leu Lys Glu Val Val Pro Gln Leu Ser Cys Pro Ile Ala Leu
130 135 140

Phe Thr
145

<210> 35
<211> 516
<212> DNA
<213> Triticum aestivum

<220>
<221> unsure
<222> (2)

<220>
<221> unsure
<222> (4)

<220>
<221> unsure
<222> (68)

<220>
<221> unsure
<222> (173)

<220>
<221> unsure
<222> (198)

<220>
<221> unsure
<222> (212)

<220>
<221> unsure
<222> (221)

<220>
<221> unsure
<222> (224)

<220>
<221> unsure
<222> (235)

<220>
<221> unsure
<222> (368)

<220>
<221> unsure
<222> (393)

<220>
<221> unsure
<222> (414)

<220>
<221> unsure
<222> (428)

<220>
<221> unsure
<222> (437)

<220>
<221> unsure
<222> (442)

<220>
<221> unsure
<222> (461)

<220>
<221> unsure
<222> (478)

<220>
<221> unsure
<222> (483)

<220>
<221> unsure
<222> (493)

<220>
<221> unsure
<222> (510)

<400> 35
angntaagcc ctccatcgtt tctctccaca gaacggattc aatgtctgga aatccagcaa 60
gccacggnct cggcggggtc gctcgccgag gcgcccgtgc ctgtcgccgc cggcgaccgc 120
ggcctgtccg tgtcgcaagc catgtcaaaa gtcatggaga agggcaagac ggncttaatc 180
ccgtacatca acgccggnga ccccgacctg gngacgacgg nggnggcgct gaggntcctc 240
gacgccctgg gcgcggacgt cgtcgagctc ggcatgccct tctcggatgc ctccgccgac 300
ggggncgtca tcaaagactc cgcggcgcgc gcgctggccg ctggcgcgac ggctgacgcc 360
atcatggnga tgctgaagga ggtganaccg cancttcct gccccgtggg gatnttctct 420
tacttcanac ccattgnaca angagggatg gggagcttcc nccccccccgt caaagaancc 480
ggnggtgaaa ggncttataa taacccgatn tttctt 516

<210> 36
<211> 159
<212> PRT
<213> Triticum aestivum

<220>
<221> UNSURE
<222> (21)

<220>
<221> UNSURE
<222> (56)

<220>
<221> UNSURE
<222> (72)..(73)

<220>
<221> UNSURE
<222> (77)

<220>
<221> UNSURE
<222> (100)

<220>
<221> UNSURE
<222> (121)

<220>
<221> UNSURE
<222> (127)

<220>
<221> UNSURE
<222> (129)

<220>
<221> UNSURE
<222> (136)

<220>
<221> UNSURE
<222> (141)

<220>
<221> UNSURE
<222> (144)

<220>
<221> UNSURE
<222> (146)

<220>
<221> UNSURE
<222> (152)

<220>
<221> UNSURE
<222> (158)

<400> 36
Ser Pro Pro Ser Phe Leu Ser Thr Glu Arg Ile Gln Cys Leu Glu Ile
1 5 10 15

Gln Gln Ala Thr Xaa Ser Ala Gly Ser Leu Ala Glu Ala Pro Val Pro
20 25 30

Val Ala Ala Gly Asp Arg Gly Leu Ser Val Ser Gln Ala Met Ser Lys
35 40 45

Val Met Glu Lys Gly Lys Thr Xaa Leu Ile Pro Tyr Ile Asn Ala Gly
50 55 60

Asp Pro Asp Leu Xaa Thr Thr Xaa Xaa Ala Leu Arg Xaa Leu Asp Ala
65 70 75 80

Leu Gly Ala Asp Val Val Glu Leu Gly Met Pro Phe Ser Asp Ala Ser
85 90 95

Ala Asp Gly Xaa Val Ile Lys Asp Ser Ala Ala Arg Ala Leu Ala Ala
100 105 110

Gly Ala Thr Ala Asp Ala Ile Met Xaa Met Leu Lys Glu Val Xaa Pro
115 120 125

Xaa Leu Ser Cys Pro Val Gly Xaa Phe Ser Tyr Phe Xaa Pro Ile Xaa
130 135 140

Gln Xaa Gly Met Gly Ser Phe Xaa Pro Pro Val Lys Glu Xaa Gly
145 150 155

<210> 37
<211> 475
<212> DNA
<213> Triticum aestivum

<400> 37
gcacgagcgg acccctacat cgacgggccc atcatccagg cgtcggtggc gcgggctctg 60
gccagcggca ccaccatgga cgccgtgctg gagatgctca gggaggtgac gccggagctg 120
tcgtgccccg tggtgctcct ctcctactac aagcccatca tgtctcgcag cttggccgag 180
atgaaagagg cgggggtcca cggtcttata gtgcctgatc tcccgtacgt ggccgcgcac 240
tcgctgtgga gtgaagccaa gaacaacaac ctggagctgg tgctgctgac aacaccagcc 300
ataccagagg acaggatgac ggagatcacc aacgcttcaa aaggcttcgt ctacctgggg 360
agcgtgaacg gagtgacagg tcctcgcgca aacgtgaacc cgccagtggg ttcacccatc 420
aagcgagtta agaaggtggc taacaagccc gttgccggtg ggcttcggga aatcc 475

<210> 38
<211> 158
<212> PRT
<213> Triticum aestivum

<400> 38
Ala Arg Ala Asp Pro Tyr Ile Asp Gly Pro Ile Ile Gln Ala Ser Val
1 5 10 15

Ala Arg Ala Leu Ala Ser Gly Thr Thr Met Asp Ala Val Leu Glu Met
20 25 30

Leu Arg Glu Val Thr Pro Glu Leu Ser Cys Pro Val Val Leu Leu Ser
35 40 45

Tyr Tyr Lys Pro Ile Met Ser Arg Ser Leu Ala Glu Met Lys Glu Ala
50 55 60

Gly Val His Gly Leu Ile Val Pro Asp Leu Pro Tyr Val Ala Ala His
65 70 75 80

Ser Leu Trp Ser Glu Ala Lys Asn Asn Asn Leu Glu Leu Val Leu Leu
85 90 95

Thr Thr Pro Ala Ile Pro Glu Asp Arg Met Thr Glu Ile Thr Asn Ala
100 105 110

Ser Lys Gly Phe Val Tyr Leu Gly Ser Val Asn Gly Val Thr Gly Pro
115 120 125

Arg Ala Asn Val Asn Pro Pro Val Gly Ser Pro Ile Lys Arg Val Lys
130 135 140

Lys Val Ala Asn Lys Pro Val Ala Gly Gly Leu Arg Glu Ile
145 150 155

<210> 39
<211> 608
<212> PRT
<213> Ruta graveolens

<400> 39
Met Ile Thr Leu Asn Val Glu Thr Pro Pro Leu Thr Arg Ser Gln Leu
1 5 10 15

Pro Ser Thr Phe Arg Val Ser Ser Ala Ala Ser Val Asn Phe Asn Asp
20 25 30

Arg Val Ala Thr Ser Arg Trp Arg Pro Asn Ser Leu Ser Leu Thr Thr
35 40 45

Ser Ser Tyr Arg Leu Arg Thr Leu Lys Cys Ala Ala Ser Ala Ser Thr
50 55 60

Ser Ala Ser Thr Ser Ala Ser Pro Ser Pro Ser Pro Ser Leu Val Asp
65 70 75 80

Gln Ser Ala Asn Phe His Glu Ala Ser Lys Lys Gly Asn Leu Ile Pro
85 90 95

Leu Tyr Arg Cys Ile Phe Ser Asp His Leu Thr Pro Val Leu Ala Tyr
100 105 110

Arg Cys Leu Val Lys Glu Asp Asp Arg Asp Ala Pro Ser Phe Leu Phe
115 120 125

Glu Ser Val Glu Pro Gly Ser Gln Ala Ser Ser Ile Gly Arg Tyr Ser
130 135 140

Val Val Gly Ala Gln Pro Ala Ile Glu Ile Val Ala Lys Glu Asn Met
145 150 155 160

Val Thr Ile Leu Asp His Glu Gly Gly Gln Arg Thr Glu Gln Phe Val
165 170 175

Glu Asp Pro Met Asp Val Pro Arg Arg Ile Met Glu Gly Trp Lys Pro
180 185 190

Gln Leu Ile Asp Glu Leu Pro Glu Ala Phe Cys Gly Gly Trp Val Gly
195 200 205

Tyr Phe Ser Tyr Asp Thr Val Arg Tyr Val Glu Lys Lys Lys Leu Pro
210 215 220

Phe Phe Ser Ala Pro Thr Asp Asp Arg Asn Leu Pro Asp Val His Leu
225 230 235 240

Gly Leu Tyr Asp Asp Val Ile Val Phe Asp His Val Glu Lys Lys Ala
245 250 255

Phe Val Ile His Trp Val Arg Leu Asp Gln Tyr Ser Ser Val Ala Glu
260 265 270

Ala Tyr Asn Asp Gly Met Asn Arg Leu Glu Asn Leu Val Ser Arg Val
275 280 285

His Asp Ile Val Pro Pro Lys Leu Arg Ser Gly Ser Ile Lys Leu His
290 295 300

Thr Arg His Phe Gly Pro Lys Leu Glu Arg Ser Ser Met Thr Ser Glu
305 310 315 320

Ala Tyr Lys Glu Ala Val Leu Glu Ala Lys Glu His Ile Leu Ala Gly
325 330 335

Asp Ile Phe Gln Ile Val Leu Ser Gln Arg Phe Glu Arg Arg Thr Phe
340 345 350

Ala Asp Pro Phe Glu Ile Tyr Arg Ser Leu Arg Ile Val Asn Pro Ser
355 360 365

Pro Tyr Met Thr Tyr Leu Gln Ala Arg Gly Cys Ile Leu Val Ala Ser
370 375 380

Ser Pro Glu Ile Leu Thr Arg Val Lys Lys Arg Lys Ile Thr Asn Arg
385 390 395 400

Pro Leu Ala Gly Thr Ile Arg Arg Gly Lys Thr Arg Lys Glu Asp Leu
405 410 415

Val Phe Glu Lys Glu Leu Leu Asn Asp Glu Lys Gln Cys Ala Glu His
420 425 430

Ile Met Leu Val Asp Leu Gly Arg Asn Asp Val Gly Lys Val Ser Glu
435 440 445

Pro Gly Ser Val Lys Val Glu Lys Leu Met Asn Ile Glu His Tyr Ser
450 455 460

His Val Met His Ile Ser Ser Thr Val Thr Gly Glu Leu Leu Asp His
465 470 475 480

Leu Thr Ser Trp Asp Ala Leu Arg Ala Ala Leu Pro Val Gly Thr Val
485 490 495

Ser Gly Ala Pro Lys Val Lys Ala Met Glu Ile Ile Asp Lys Leu Glu
500 505 510

Val Thr Arg Arg Gly Pro Tyr Gly Gly Gly Phe Gly Gly Ile Ser Phe
515 520 525

Thr Gly Asp Leu Asp Ile Ala Leu Ala Leu Arg Thr Met Val Phe Gln
530 535 540

Thr Ala Thr Arg Tyr Asp Thr Met Tyr Ser Tyr Lys Asp Val Asp Lys
545 550 555 560

Arg Arg Glu Trp Ile Ala His Leu Gln Ala Gly Ala Gly Ile Val Ala
565 570 575

Asp Ser Asp Pro Ala Asp Glu Gln Arg Glu Cys Glu Asn Lys Ala Ala
580 585 590

Ala Leu Ala Arg Ala Ile Asp Leu Ala Glu Ser Ser Phe Ile Glu Lys
595 600 605

<210> 40
<211> 276
<212> PRT
<213> Arabidopsis thaliana

<400> 40
Met Ala Ala Ser Thr Leu Tyr Lys Ser Cys Leu Leu Gln Pro Lys Ser
1 5 10 15

Gly Ser Thr Thr Arg Arg Leu Asn Pro Ser Leu Val Asn Pro Leu Thr
20 25 30

Asn Pro Thr Arg Val Ser Val Leu Gly Lys Ser Arg Arg Asp Val Phe
35 40 45

Ala Lys Ala Ser Ile Glu Met Ala Glu Ser Asn Ser Ile Pro Ser Val
50 55 60

Val Val Asn Ser Ser Lys Gln His Gly Pro Ile Ile Val Ile Asp Asn
65 70 75 80

Tyr Asp Ser Phe Thr Tyr Asn Leu Cys Gln Tyr Met Gly Glu Leu Gly
85 90 95

Cys His Phe Glu Val Tyr Arg Asn Asp Glu Leu Thr Val Glu Glu Leu
100 105 110

Lys Lys Lys Asn Pro Arg Gly Val Leu Ile Ser Pro Gly Pro Gly Thr
115 120 125

Pro Gln Asp Ser Gly Ile Ser Leu Gln Thr Val Leu Glu Leu Gly Pro
130 135 140

Leu Val Pro Leu Phe Gly Val Cys Met Gly Leu Gln Cys Ile Gly Glu
145 150 155 160

Ala Phe Gly Gly Lys Ile Val Arg Ser Pro Phe Gly Val Met His Gly
165 170 175

Lys Ser Ser Met Val His Tyr Asp Glu Lys Gly Glu Glu Gly Leu Phe
180 185 190

Ser Gly Leu Ser Asn Pro Phe Ile Val Gly Arg Tyr His Ser Leu Val
195 200 205

```
Ile Glu Lys Asp Thr Phe Pro Ser Asp Glu Leu Glu Val Thr Ala Trp
    210                 215                 220

Thr Glu Asp Gly Leu Val Met Ala Ala Arg His Arg Lys Tyr Lys His
225                 230                 235                 240

Ile Gln Gly Val Gln Phe His Pro Glu Ser Ile Ile Thr Thr Glu Gly
                245                 250                 255

Lys Thr Ile Val Arg Asn Phe Ile Lys Ile Val Glu Lys Lys Glu Ser
            260                 265                 270

Glu Lys Leu Thr
        275
```

<210> 41
<211> 312
<212> PRT
<213> Arabidopsis thaliana

<400> 41

```
Met Ala Ile Ala Phe Lys Ser Gly Val Phe Phe Leu Gln Ser Pro Lys
  1               5                  10                  15

Ser Gln Ile Gly Phe Arg His Ser Ser Pro Pro Asp Ser Ser Leu Ser
            20                  25                  30

Phe Lys Arg Phe Thr Pro Met Ala Ser Leu Ser Thr Ser Ser Pro Thr
        35                  40                  45

Leu Gly Leu Ala Asp Thr Phe Thr Gln Leu Lys Lys Gln Gly Lys Val
    50                  55                  60

Ala Phe Ile Pro Tyr Ile Thr Ala Gly Asp Pro Asp Leu Ser Thr Thr
65                  70                  75                  80

Ala Glu Ala Leu Lys Val Leu Asp Ala Cys Gly Ser Asp Ile Ile Glu
                85                  90                  95

Leu Gly Val Pro Tyr Ser Asp Pro Leu Ala Asp Gly Pro Val Ile Gln
            100                 105                 110

Ala Ala Ala Thr Arg Ser Leu Glu Arg Gly Thr Asn Leu Asp Ser Ile
        115                 120                 125

Leu Glu Met Leu Asp Lys Val Val Pro Gln Ile Ser Cys Pro Ile Ser
    130                 135                 140

Leu Phe Thr Tyr Tyr Asn Pro Ile Leu Lys Arg Gly Leu Gly Lys Phe
145                 150                 155                 160

Met Ser Ser Ile Arg Ala Val Gly Val Gln Gly Leu Val Val Pro Asp
                165                 170                 175

Val Pro Leu Glu Glu Thr Glu Met Leu Arg Lys Glu Ala Leu Asn Asn
            180                 185                 190

Asp Ile Glu Leu Val Leu Leu Thr Thr Pro Thr Thr Pro Thr Glu Arg
        195                 200                 205

Met Lys Leu Ile Val Asp Ala Ser Glu Gly Phe Ile Tyr Leu Val Ser
    210                 215                 220

Ser Ile Gly Val Thr Gly Ala Arg Ser Ser Val Ser Gly Lys Val Gln
225                 230                 235                 240
```

```
Ser Leu Leu Lys Asp Ile Lys Glu Ala Thr Asp Lys Pro Val Ala Val
                245                 250                 255

Gly Phe Gly Ile Ser Lys Pro Glu His Val Lys Gln Ile Ala Gly Trp
            260                 265                 270

Gly Ala Asp Gly Val Ile Val Gly Ser Ala Met Val Lys Leu Leu Gly
        275                 280                 285

Asp Ala Lys Ser Pro Thr Glu Gly Leu Lys Glu Leu Glu Lys Leu Thr
    290                 295                 300

Lys Ser Leu Lys Ser Ala Leu Leu
305                 310
```

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/60, 9/88, 1/21, C12Q 1/68

A3

(11) International Publication Number: **WO 99/49058**

(43) International Publication Date: 30 September 1999 (30.09.99)

(21) International Application Number: PCT/US99/06046

(22) International Filing Date: 19 March 1999 (19.03.99)

(30) Priority Data:
60/079,386 26 March 1998 (26.03.98) US

(71) Applicant *(for all designated States except US)*: E.I. DU PONT DE NEMOURS AND COMPANY [US/US]; 1007 Market Street, Wilmington, DE 19898 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: VOLLMER, Steven, J. [US/US]; 1019 Tweedbrook Road, Wilmington, DE 19810 (US). FALCO, Saverio, Carl [US/US]; 1902 Millers Road, Arden, DE 19810 (US). BROGLIE, Richard, M. [US/US]; 520 Port Royal Court, Landenberg, PA 19350 (US). BRYAN, Gregory, T. [NZ/US]; 1215 Spruce Street, Wilmington, DE 19805 (US). CAHOON, Rebecca, E. [US/US]; 2331 West 18th Street, Wilmington, DE 19806 (US). RAFALSKI, J., Antoni [US/US]; 2028 Longcome Drive, Wilmington, DE 19810 (US).

(74) Agent: MAJARIAN, William, R.; E.I. du Pont de Nemours and Company, Legal Patent Records Center, 1007 Market Street, Wilmington, DE 19898 (US).

(81) Designated States: AL, AU, BA, BB, BG, BR, CA, CN, CU, CZ, EE, GD, GE, HR, HU, ID, IL, IN, IS, JP, KP, KR, LC, LK, LR, LT, LV, MG, MK, MN, MX, NO, NZ, PL, RO, SG, SI, SK, SL, TR, TT, UA, US, UZ, VN, YU, ZA, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(88) Date of publication of the international search report:
6 April 2000 (06.04.00)

(54) Title: TRYPTOPHAN BIOSYNTHETIC ENZYMES

(57) Abstract

This invention relates to an isolated nucleic acid fragment encoding a tryptophan biosynthetic enzyme. The invention also relates to the construction of a chimeric gene encoding all or a portion of the tryptophan biosynthetic enzyme, in sense or antisense orientation, wherein expression of the chimeric gene results in production of altered levels of the tryptophan biosynthetic enzyme in a transformed host cell.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

INTERNATIONAL SEARCH REPORT

Inte onal Application No
PCT/US 99/06046

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/60 C12N9/88 C12N1/21 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category '	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 97 26366 A (DEKALB GENETICS CORP ;ANDERSON PAUL C (US); CHOMET PAUL S (US); GR) 24 July 1997 (1997-07-24) the whole document --- -/--	1-34, 54-59

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

' Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
18 January 2000	03.02.00

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl. Fax: (+31-70) 340-3016	Smalt, R

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 99/06046

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication where appropriate, of the relevant passages	Relevant to claim No
X	DATABASE EMBL - EMEST20 'Online! Entry ZM248. Acc.no.T25248. 1 October 1994 (1994-10-01) SHEN, B. ET AL.: "5c10h08 membrane-free polysomes from endosperm Zea mays cDNA clone 5c10h08 5' end similar to anthranilate synthase." XP002112345 the whole document -& SHEN, B. ET AL.: "Partial sequencing and mapping of clones from two maize cDNA libraries." PLANT MOLECULAR BIOLOGY, vol. 26, 1994, pages 1085-1101, XP002042536 the whole document	1,2,10, 18,22, 23,28
X	BOHLMANN J ET AL: "PURIFICATION AND CDNA CLONING OF ANTHRANILATE SYNTHASE FROM RUTA GRAVEOLENS, MODES OF EXPRESSION AND PROPERTIES OF NATIVE AND RECOMBINANT ENZYMES" PLANT JOURNAL,GB,BLACKWELL SCIENTIFIC PUBLICATIONS, OXFORD, vol. 7, no. 3, page 491-501 XP002032323 ISSN: 0960-7412 the whole document	1,5,6, 13,14
X	TOZAWA, Y. ET AL.: "Isolation and analysis of gene coding for alpha-subunit of rice anthranilate synthase. (in Japanese)" JAPANESE JOURNAL OF BREEDING, vol. 46, no. 2, 27 September 1996 (1996-09-27), page 28 XP002917196 see whole document; cited in search report of WO9911800 (first citation). the whole document -& DATABASE EMBL - TREMBL 'Online! Entry/acc.no. Q9XJ29, 1 November 1999 (1999-11-01) TOZAWA, Y. ET AL.: "Anthranilate synthase alpha 2 subunit." XP002127895 the whole document	4,9,11, 18

-/--

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Inte onal Application No
PCT/US 99/06046

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication where appropriate, of the relevant passages	Relevant to claim No
X	DATABASE EMBL - EMPLN 'Online! Entry ZMORF, Acc.no.M95067, 9 June 1992 (1992-06-09) KEITH, C.S. ET AL.: "Zea mays putative anthranilate synthase component II homolog mRNA, partial cds" XP002112490 the whole document -& KEITH, C.S. ET AL.: "Partial sequence analysis of 130 randomly selected maize cDNA clones." PLANT PHYSIOLOGY, vol. 101, 1993, pages 329-32, XP002112488 the whole document ---	22-25,27
X	NIYOGI K K ET AL: "SUPPRRESSORS OF TRP1 FLUORESCENCE IDENTIFY A NEW ARABIDOPSIS GENE, TRP4, ENCODING THE ANTHRANILATE SYNTHASE BETA SUBUNIT" PLANT CELL, vol. 5, 1 September 1993 (1993-09-01), pages 1011-1027, XP002032325 ISSN: 1040-4651 abstract; figures 1,7,8 ---	1-4, 10-12, 22-26, 29-34
X	DATABASE EMBL - EMEST20 'Online! Entry ZM7731, Acc.no.T8773, 14 May 1994 (1994-05-14) SHEN, B. ET AL.: "2C06H01-T7 membrane-free polysomes from endosperm Zea mays cDNA clone 2C06H01 5' end similar to glucocorticoid receptor." XP002112491 the whole document -& SHEN, B. ET AL.: "Partial sequencing and mapping of clones from two maize cDNA libraries." PLANT MOLECULAR BIOLOGY, vol. 26, November 1994 (1994-11), pages 1085-1101, XP002042536 the whole document ---	22,23,27
X	DATABASE EMBL - EMEST3 'Online! Entry/Acc.No. AA751371, 21 January 1998 (1998-01-21) NAHM, B.H. ET AL.: "ISYS0037 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa cDNA clone ISYS0037 similar to MZEORFH Zea mays putative anthranilate synthase component II homolog mRNA, partial cds." XP002112492 the whole document --- -/--	22,24,28

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 99/06046

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No
X	RADWANSKI, E. R. ET AL.: "Arabidopsis thalia tryptophan synthase alpha: gene cloning, expression, and subunit interaction." MOLECULAR AND GENERAL GENETICS. vol. 248, 1995, pages 657-67, XP002127894 the whole document	35-53
X	KRAMER V C ET AL: "STRUCTURE OF A MAIZE TRYPTOPHAN SYNTHASE ALPHA SUBUNIT GENE WITH PITH ENHANCED EXPRESSION" PLANT MOLECULAR BIOLOGY,NL,NIJHOFF PUBLISHERS, DORDRECHT, vol. 27, 1995, pages 1183-1188-8, XP002071677 ISSN: 0167-4412	35-37, 39-42
Y	abstract; figure 1	38
Y	DATABASE EMBL - EMEST22 'Online! Entry/Acc.no. C27627, 6 August 1997 (1997-08-06) YAMAMOTO, K. ET AL.: "Rice cDNA, partial sequence." XP002127896 the whole document	38
X	DATABASE EMBL - EMEST2 'Online! Entry/Acc.no. AA660642, 14 November 1997 (1997-11-14) COVITZ, P.A. ET AL.: "00530 MtRHE Medicago truncatula cDNA 5' similar to tryptophan synthase alpha chain." XP002127897 the whole document	35,40
P,X	WO 99 11800 A (HASEGAWA HISAKAZU ;HOKKO CHEM IND CO (JP); TERAKAWA TERUHIKO (JP);) 11 March 1999 (1999-03-11) page 125 -page 131; claims 1,2 -& DATABASE WPI Section Ch, Week 9919 Derwent Publications Ltd., London, GB; Class C06, AN 99-228982 XP002112346 the whole document	1-13, 18-21, 54-58

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 99/06046

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [ ] Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. [X] Claims Nos.: 57,58
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
see FURTHER INFORMATION sheet PCT/ISA/210

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. [X] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[X] No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US 99/06046

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 57,58

Present claims 57 and 58, relating to products which can be obtained by the methods of claims 55 and 56, respectively, have not been searched due to lack of support for such products within the meaning of Article 6 PCT and/or lack of disclosure of such products within the meaning of Article 5 PCT in the description.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US 99/06046

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 2,3,10,11,23,24,27,28 completely, and 1,18-22, 31-34,54-59 partially

   Corn anthranilate synthase subunits alpha and beta, portions thereof, nucleic acids encoding them and sequences with at least 80% homology thereto, chimeric genes comprising a portion of said nucleic acids, hosts transformed with said nucleic acids, method for altering expression of said gene, method for obtaining said nucleic acids, and method for evaluating the ability of a compound to inhibit the activity of said protein.

2. Claims: 36,37,43,44 completely, and 35,50-59 partially

   As invention 1, but limited to corn tryptophan synthase subunit alpha.

3. Claims: 4,12,25,29 completely, and 1,18-22, 31-34, 54-59 partially

   As invention 1, but limited to rice anthranylate synthase subunits alpha and beta.

4. Claims: 38,39,45,46 completely, and 35,50-59 partially

   As invention 1, but limited to rice tryptophan synthase subunit alpha.

5. Claims: 5,6,13,14,40,47 completely, and 1,18-21,35, 50-59 partially

   As invention 1, but limited to soybean anthranylate synthase subunit alpha and tryptophane synthase subunit alpha.

6. Claims: 7-9,15-17,26,30,41,42,48,49 completely, and 1, 18-22,31-35,50-59 partially

   As invention 1, but limited to wheat anthranylate synthase subunits alpha and beta, and tryptophane synthase subunit alpha.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 99/06046

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9726366 A	24-07-1997	AU 1706697 A	11-08-1997
		BR 9707055 A	20-07-1999
		CA 2243268 A	24-07-1997
		EP 0879292 A	25-11-1998
		HU 9900387 A	28-05-1999
WO 9911800 A	11-03-1999	AU 8888098 A	22-03-1999